

## 明 細 書

## 楽曲再生方法及び装置並びに音楽音響データ中のサビ区間検出方法

## 技術分野

本発明は、楽曲中のサビ区間等の特徴区間を選択的に再生する方法及び装置並びに特徴区間（サビ区間）を検出する方法に関するものである。

## 背景技術

従来、音楽情報処理の研究分野では楽曲の検索や音楽理解に関する研究は多かったものの、音楽の試聴に着目した研究はなされていなかった。現在、レコード店の店頭でコンパクトディスク（CD）等に記録された音楽を「試聴」する場合、試聴者は、早送りを行いながら聴きたい部分だけを選び出して試聴している場合が多い。これは試聴の主目的が、自分の探していた楽曲、好みの楽曲であるかどうかを短時間で判断するためのものだからである。例えばポピュラー音楽の場合等では、楽曲中で一番代表的な盛り上がる主題の部分であるサビ（chorus, refrain）区間や何度も繰り返し演奏されるメロディ区間のような楽曲の構造に特徴のある区間（以下楽曲構造特徴区間と言う）を試聴して判断することが多い。そこで試聴者は、イントロを少し聴いた後に、サビ区間や繰り返し区間等の楽曲構造特徴区間を探しながら早送りボタンを何度も押して途中を飛ばし、楽曲構造特徴区間を再生するというような特殊な聴き方をする。

しかし、従来の音楽CDの試聴機には、このような試聴固有の聴き方を支援する機能はなかった。試聴機は通常のCDプレーヤ相当の再生操作ボタンを持つが、その中で、早送りと早戻しのボタンしか、サビの部分を探すために利用できなかった。一方、最近CD販売店に導入され始めたデジタル試聴機では、MP3等の圧縮形式で蓄積されている数十万曲の中から、ハードディスクあるいはネットワーク経由で再生することができる。しかし、楽曲先頭の短い区間（通常45秒）だけが機械的に切り出されて収録されているため、試聴者は楽曲構造特徴部分を必ずしも聴くことはできなかった。近年、日本のポピュラー音楽ではサビから

始まる楽曲構造を持つ曲が増えているとはいえ、発明者の調査では、日本のポピュラー音楽のヒットチャート（2001年1月～12月の週間ランキングのシングル上位20曲）の楽曲中、楽曲開始後40秒以内にサビが始まる曲は約20%しかなかった。

従来のサビ検出方法の一つでは、楽曲の音響信号の代表的部分として、サビを指定した長さだけ不完全に一箇所切り出す。Logan等〔先行技術1〕は、切り出した短いフレーム（1秒間）にその部分の特徴量に基づいてラベルを付与し、最頻出のラベルをもつフレームをサビとみなす方法を提案した。このラベルの付与には、各区間の特徴量間の類似度に基づくクラスタリングや隠れマルコフモデルを用いていた。また、Bartsch等〔先行技術2〕は、ビートトラッキングの結果に基づいて楽曲を拍ごとの短いフレームに分割し、それらの特徴量間の類似度が、指定した一定の長さの区間に渡って最も高い箇所を、サビとして切り出す方法を提案した。また、Foote〔先行技術3〕は、非常に短い断片（フレーム）ごとの特徴量間の類似度に基づく境界検出の応用例として、サビが切り出せる可能性を指摘していた。

一方、標準MIDIファイル等の音符相当表現を対象とした従来技術〔先行技術4および5〕もあるが、この技術は音源分離が困難な混合音にはそのまま適用できなかった。また従来のサビ区間を検出する技術では、常に指定した一定の長さを切り出して提示するだけで、サビの区間がどこからどこまでかは推定していなかった。さらに従来の技術では、いずれも転調を考慮していなかった。

なお先行する技術としては以下のものがある。

先行技術1: Logan, B. and Chu, S.: Music Summarization Using Key Phrases, Proc. of ICASSP 2000, II-749-752 (2000).

先行技術2: Bartsch, M. A. and Wakefield, G. H.: To Catch A Chorus: Using Chroma-based Representations for Audio Thumbnailing, Proc. of WASPAA 2001, 15-18 (2001).

先行技術3: Foote, J. : Automatic Audio Segmentation Using A Measure of Audio Novelty, Proc. of ICME 2000, 1-452-455 (2000).

先行技術4: Meek, C. and Birmingham, W. P. : Thematic Extractor, Proc. of ISMIR 2001, 119-128 (2001).

先行技術5: 村松純: 歌謡曲における「さび」の楽譜情報に基づく特徴抽出—小室哲哉の場合—, 情処研報 音楽情報科学, 2000-MUS-35-1, 1-6 (2000).

先行技術6: 大津展之: 判別および最小2乗規準に基づく自動しきい値選定法, 信学論(D), J63-D, 4, 349-356 (1980).

先行技術7: Shepard, R. N. : Circularity in Judgments of Relative Pitch, J. Acoust. Soc. Am., 36, 12, 2346-2353 (1964).

先行技術8: Wakefield, G. H. : Mathematical Representation of Joint Time-Chroma Distributions, SPIE 1999, 637-645 (1999).

先行技術9: Savitzky, A. and Golay, M. J. : Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Analytical Chemistry, 36, 8, 1627-1639 (1964).

先行技術10: 後藤真孝, 橋口博樹, 西村拓一, 岡隆一: RWC研究用音楽データベース; ポピュラー音楽データベースと著作権切れ音楽データベース, 情処研報 音楽情報科学, 2001-MUS-42-6, 35-42 (2001).

先行技術11: van Rijsbergen, C. J. : Information Retrieval, Butterworths, second ed

ition (1979) .

先行技術 12 : 平田圭二, 松田周 : パピプーン : GTTMに基づく音楽要約システム, 情報研報 音楽情報科学, 2002-MUS-46-5, 29-36 (2002) .

本発明の目的は、インターフェースを用いた選択により簡単に楽曲構造特徴区間を、音楽音響データ再生装置を用いて再生することができる楽曲再生方法及び装置並びに該装置に用いるインターフェース及びプログラムを提供することにある。

本発明の他の目的は、特に楽曲中のサビ区間を音楽音響データ再生装置を用いて簡単に再生することができる楽曲再生方法及び装置並びに該装置に用いるインターフェースを提供することにある。

本発明の他の目的は、楽曲中のサビ区間の特定を確実にできる楽曲再生方法及び装置並びに該装置に用いるインターフェースを提供することにある。

本発明の別の目的は、楽曲構造特徴区間の分散状況と音楽音響データの再生状況とを視覚により確認できる楽曲再生方法及び装置並びに該装置に用いるインターフェースを提供することにある。

本発明の他の目的は、サビ出し区間と繰り返し区間の存在状況を視覚により区別することができるようにした楽曲再生方法及び装置並びに該装置に用いるインターフェースを提供することにある。

本発明の他の目的は、操作者が選択ボタンを操作するだけで楽曲構造特徴区間を選択的に再生することができる楽曲再生方法及び装置並びに該装置に用いるインターフェースを提供することにある。

本発明の別の目的は、楽曲構造特徴区間を統計データから容易に抽出することができる楽曲構造特徴区間の抽出方法を提供することにある。

本発明の他の目的は、従来の技術の問題点を克服し、楽曲中に出現するサビ区間を網羅的に検出することができる音楽音響データ中のサビ区間を検出する方法及び装置並びにプログラムを提供することにある。

本発明の更に他の目的は、1つのサビ区間がどこからどこまでかを検出することができる音楽音響データ中のサビ区間を検出する方法及び装置並びにプログラ



ムを提供することにある。

本発明の他の目的は、転調されたサビ区間も検出できる音楽音響データ中のサビ区間を検出する方法及び装置並びにプログラムを提供することにある。

本発明の他の目的は、サビ区間だけでなく、その他の繰り返し区間も表示手段に表示することができる音楽音響データ中のサビ区間を検出する装置を提供することにある。

本発明のさらに他の目的は、サビ区間だけでなく、その他の繰り返し区間も再生することができる音楽音響データ中のサビ区間を検出する装置を提供することにある。

#### 発明の開示

サビは、楽曲全体の構造の中で、一番代表的な盛り上がる主題の部分である。通常、サビは楽曲中で最も多く繰り返され、印象に残るため、専門的な音楽の訓練を受けていない人が音楽を聴いたときでも、どこがサビであるかを容易に判断できる。さらに、サビ検出の結果は、様々な応用において有用である。例えば、多数の楽曲をブラウジングするときや、楽曲検索システムにおいて検索結果を提示するときに、サビの冒頭を短く再生（プレビュー）できると便利である（画像のサムネールの音楽版とみなせる）。また、歌声等を検索キーとした楽曲検索では、検索対象をサビ区間に限定すると精度と効率が上がる。そして本発明のサビ検出技術を実施すれば、サビ区間を自動的にインデキシングすることも可能になる。

音楽音響データを再生する音楽音響データ再生装置を用いて音楽音響データを再生する本発明の楽曲再生方法は、第1乃至第3のステップを実行する。ここで音楽音響データとは、アナログ音響信号、デジタル音響信号、MIDIデータ（標準MIDIファイル）等の各種の再生可能な音響データを含むもので、圧縮／非圧縮のいずれでもよい。なお圧縮されたデータの場合、圧縮フォーマットの形式は任意である。また後述する「複数の楽曲構造特徴区間」とは、前述のサビ区間及び／または繰り返し区間を含むものである。なおサビ区間または繰り返し区間だけを複数の楽曲構造特徴区間としてもよいのは勿論である。

第1のステップでは、再生の対象となる音楽音響データに関して、複数の楽曲構造特徴区間を指定するための複数の楽曲構造特徴区間指定データを予め用意する。複数の楽曲構造特徴区間指定データとは、楽曲構造特徴区間の始まり位置と終了位置に関するデータである。例えば、楽曲構造特徴区間の始まり位置と終了位置は、対象となる音楽音響データ中の位置データや、実際に演奏を開始してからその位置に達するまでに要する時間で表した位置データとして表現することができる。この複数の楽曲構造特徴区間指定データは、音楽音響データと組合せてすなわちセットにしてもよいが、音楽音響データとは別のデータとして存在させてもよい。このようにすると、すでに音楽音響データを入手している人は、ネットワーク等を介してその楽曲に対応する複数の楽曲構造特徴区間指定データを後から入手することができるようになる。また複数の楽曲構造特徴区間指定データを音楽音響データと組合せておけば、音楽音響データをダウンロードするときに必然的に複数の楽曲構造特徴区間指定データもダウンロードされる上、両データの対応関係を考慮する必要がなくなるので便利である。

複数の楽曲構造特徴区間の抽出方法は、任意である。例えば、音楽音響データから任意の方法で自動抽出してもよい。あるいは、音楽音響データを再生しながら複数の楽曲構造特徴区間を人が特定したり（ラベリングし）、専用のラベリング用エディタを用いて複数の楽曲構造特徴区間を特定してもよい。また対象とする音楽音響データを複数の試聴者が再生した際の再生行動の統計データに基いて複数の楽曲構造特徴区間を特定し、特定した複数の楽曲構造特徴区間を指定するための複数の楽曲構造特徴区間指定データを作成するようにしてもよい。この場合には、例えば、複数の試聴者が試聴した回数または再生した回数が多い区間（予め定めた回数よりも多い区間）を楽曲構造特徴区間として特定することができる。なおこの場合には、ネットワークに接続された複数台の試聴用の楽曲再生装置から複数の試聴者による再生回数が多い区間のデータを集めてもよい。このようにすると簡単に大量のデータを入手できる。また入手したデータをヒストグラムとして視覚により認識可能な状態とし、ヒストグラムから複数の楽曲構造特徴区間を特定してもよい。このようにすると統計データからの楽曲構造特徴区間の特定が容易になる。このような人の試聴を基準にすると、楽曲構造からサビとい

われる部分の検出が困難な楽曲からも楽曲構造特徴部分を容易に抽出することができる。

第2のステップでは、複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するインターフェースを用意する。このインターフェースとは、操作者と装置の間を関係づけるものであり、操作者により選択操作される部分と、選択結果を発信する機能とを少なくとも備えていけばよい。したがって最も簡単なインターフェースは、アナログまたはデジタルのスイッチからなる選択ボタンにより構成することができる。選択ボタンだけでインターフェースを構成する場合には、再生される音楽を聴きながら選択ボタンを操作することになる。選択ボタンだけでは、音楽音響データのどの部分に楽曲構造特徴区間が幾つ存在するのかが分からない。そのため必要以上に選択ボタンを操作する可能性がある。また選択ボタンだけでは、複数種類の楽曲構造特徴区間がある場合に同じ種類の楽曲構造特徴区間を意図的に試聴することができない。

そこでこのような問題を解決するためには、第2のステップで用意するインターフェースとして、音楽音響データの長さを時間軸を用いた画像表示により表示し、併せて音楽音響データ再生装置が音楽音響データを再生している音楽音響データの再生位置と複数の楽曲構造特徴区間とを時間軸を用いた画像表示により表示する画像表示部を有するものを用いければよい。このようなインターフェースを用いると、画像表示部の表示を見ることにより、現在の再生位置と複数の楽曲構造特徴区間との位置関係または時間関係を視覚を用いて認識できる。したがって画像表示部の表示を見ながら、任意の箇所の楽曲構造特徴区間を任意に選択することが可能になる。なお楽曲の演奏時間が長い場合には、演奏時間の全時間分の表示を一画面に表示することができない場合もある。このような場合には、演奏時間の一部に対応する表示を画面に表示しておき、時間の経過とともにスクロール式に画面を変更してもよいのは勿論である。

インターフェースに画像表示部を設ける場合に、複数の楽曲構造特徴区間において、楽曲のサビ部分を示す複数のサビ区間と、繰り返し部分を示す複数の繰り返し区間とから構成するときには、画像表示部が複数のサビ区間と複数の繰り返し区間とを区別して表示するのが好ましい。このようにすると画像表示を見なが

らサビ区間と繰り返し区間とを任意に選択することが容易になる。

またインターフェースは、複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するために操作者が操作する 1 以上の選択ボタンを有している。この選択ボタンとして、人が指で操作するアナログ式の実選択ボタン、表示画面に表示されたボタンの表示を人間の指やカーソルでクリックしてオン・オフを判定するタイプの選択ボタン（画像表示ボタン）、音声認識により選択を判別するタイプの選択ボタン等の各種の実選択ボタンが含まれる。特に、1 以上の実選択ボタンに、複数の楽曲構造特徴区間からサビ区間だけを選択する 1 以上のサビ区間実選択ボタンをインターフェースに配置するのが好ましい。このような専用実選択ボタンを配置すると、特別な知識を持たない人でも簡単にサビ区間の再生を実行できるので、汎用性が非常に高くなる。また 1 以上のサビ区間実選択ボタンは、前のサビ区間を選択する第 1 のサビ区間実選択ボタンと、後のサビ区間を選択する第 2 のサビ区間実選択ボタンとから構成するのが好ましい。このようにすると選択の自由度が増すだけでなく、繰り返しサビ区間を再生することが可能になる。また 1 以上の実選択ボタンには、複数の繰り返し区間を選択する際に前の繰り返し区間を選択する第 1 の繰り返し区間実選択ボタンと、後の繰り返し区間を選択する第 2 の繰り返し区間実選択ボタンが含まれてもよい。この場合にも選択の自由度が増すだけでなく、繰り返し区間を繰り返し再生することが可能になる。

第 3 のステップでは、音楽音響データ再生装置を用いて音楽音響データの再生を開始した後、操作者がインターフェースを用いて選択した任意の楽曲構造特徴区間に対応する楽曲構造特徴区間指定データに基づいて、音楽音響データ再生装置による音楽音響データの再生位置を前記任意の楽曲構造特徴区間に変更する。すなわち演奏途中であっても、インターフェースから特定の楽曲構造特徴区間に再生位置を変更する指令が入力されると、その指令に基づいて再生位置を変更することになる。

なおインターフェース上で任意の楽曲構造特徴区間を選択した場合にのみ、その選択された区間の再生を行うようにしてもよいのは勿論である。その場合には、第 3 のステップでは、操作者がインターフェースを用いて選択した任意の楽曲構造特徴区間に対応する楽曲構造特徴区間指定データに基づいて音楽音響データ再

生装置に前記任意の楽曲構造特徴区間を再生させるようにすればよい。

本発明の方法を実施するための楽曲再生装置は、音楽音響データを再生する音楽音響データ再生装置を有し、且つ楽曲中の複数の楽曲構造特徴区間から選択した任意の楽曲構造特徴区間を再生する機能を備えている。そして本発明の楽曲再生装置は、インターフェースと、楽曲構造特徴区間指定データ格納手段と、指定データ付与手段とを更に備えている。インターフェースは、複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択する機能を有するものである。また楽曲構造特徴区間指定データ格納手段は、再生する音楽音響データに対応して予め定められた複数の楽曲構造特徴区間を指定するための楽曲構造特徴区間指定データを格納する。そして指定データ付与手段は、インターフェースによって選択された任意の楽曲構造特徴区間を指定する楽曲構造特徴区間指定データを音楽音響データ再生装置に与える。音楽音響データ再生装置は、指定データ付与手段から付与された楽曲構造特徴区間指定データによって指定された楽曲構造特徴区間に再生位置を変更するように構成されている。本発明の装置では、楽曲の再生中であっても、インターフェースによって選択された任意の楽曲構造特徴区間を再生することができる。また楽曲の再生中でなくても、インターフェースによって選択された任意の楽曲構造特徴区間を再生することができる。なお音楽音響データに楽曲構造特徴区間指定データが付随している場合には、楽曲構造特徴区間指定データ格納手段は音楽音響データに付随する楽曲構造特徴区間指定データを読み取って格納すればよい。

また音楽音響データから自動的に楽曲中の複数の楽曲構造特徴区間を抽出して楽曲構造特徴区間指定データを作成する特徴区間指定データ生成手段を更に備えていてもよい。このような手段を備えていれば、既存の楽曲を再生する場合にも本発明を適用することが容易になる。

なお本発明は、コンピュータと、前記コンピュータからの指令に基いて動作するディスプレイと、前記コンピュータからの指令で音楽音響データを再生する音楽音響データ再生装置とから構成された装置を用いて、楽曲中の複数の楽曲構造特徴区間から選択した任意の楽曲構造特徴区間を再生する機能を前記コンピュータに実現させるためのプログラムとしても特定できる。このプログラムは、複数

の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するインターフェースを前記ディスプレイ上に構成する機能と、再生する前記音楽音響データに対応して予め定められた前記複数の楽曲構造特徴区間を指定するための楽曲構造特徴区間指定データを格納する楽曲構造特徴区間指定データ格納手段を構成する機能と、前記インターフェースによって選択された前記任意の楽曲構造特徴区間を指定する前記楽曲構造特徴区間指定データを前記音楽音響データ再生装置に与える指定データ付与手段を構成する機能と、前記指定データ付与手段から付与された前記楽曲構造特徴区間指定データによって指定された前記楽曲構造特徴区間を再生するように前記音楽音響データ再生装置に指令を与える機能とを前記コンピュータに実現させる構成を備えている。

なおインターフェースを前記ディスプレイ上に構成する機能は、複数の楽曲構造特徴区間に含まれるサビ区間を選択するために操作者が操作するサビ区間選択ボタンを含む1以上の選択ボタンをディスプレイ上に形成する機能を有しているのが好ましい。

本発明を用いれば、「サビ出し」機能（サビ区間選択ボタン）を搭載した音楽試聴機を得ることができる。試聴者はこのサビ区間選択ボタンを押すだけで、サビ区間の先頭へ再生をジャンプさせる（瞬時に早送りする）ことができ、自分でサビ区間を探す煩わしい作業から解放される。さらに、サビ以外の楽曲中の繰り返し区間も事前に推定あるいは用意しておくことで、次の楽曲構造の繰り返し区間の先頭へも再生をジャンプさせることができる。

また本発明のサビ区間の検出方法では、ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出するために、特徴量抽出ステップと、類似度演算ステップと、繰り返し区間リストアップステップと、統合繰り返し区間決定ステップと、サビ区間決定ステップとを実行する。

最初に、特徴量抽出ステップでは、音楽音響データから所定の時間単位で音響特徴量を順次求める。具体的な実施の形態では、入力されてくる音楽音響データについて、所定のサンプリング幅を持って重複しながらデータのサンプリングを行うハニング窓等のサンプリング技術を用いて、所定の時間単位（例えば80m

s) で、サンプリングを行う。そしてサンプリングしたデータについて、音響特徴量を求める。音響特徴量の求め方は任意である。例えば、特徴量抽出ステップで求める音響特徴量として、1 オクターブの範囲に含まれる 12 の音名の周波数のパワーを複数のオクターブに渡ってそれぞれ加算して得た 12 次元クロマベクトルを用いることができる。12 次元クロマベクトルを音響特徴量として用いると、複数オクターブに渡る楽曲の特徴量を抽出できるだけでなく、転調した音楽音響データから対比が可能な特徴量として抽出することができる。

次に、類似度演算ステップでは、音楽音響データについて求めた複数の音響特徴量の相互間の類似度を求める。類似度を求める際に用いる演算式は、任意であり、公知の類似度演算式のいずれを用いてもよい。そして繰り返し区間リストアップステップでは、類似度に基づいて音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする。類似度演算ステップで、今回求めた音響特徴量と先に求めた全ての音響特徴量との間の類似度を求めると、リアルタイムにサビ区間を検出することが可能になる。

より具体的な、類似度演算ステップでは、時刻  $t$  のクロマベクトル（音響特徴量）とそれよりラグ  $l$  ( $0 \leq l \leq t$ )（ $l$  はアルファベット  $L$  の小文字）だけ過去の全てのクロマベクトルとの類似度を求めることになる。この場合、繰り返し区間リストアップステップでは、一方の軸を時間軸とし他方の軸をラグ軸とし、予め定めた時間長さ以上類似度が予め定めた閾値以上ある場合には類似度が予め定めた閾値以上である部分の長さに対応する時間長さを有する類似線分を時間軸を基準にした繰り返し区間としてリストアップする。なおこのリストアップは、演算上のリストアップであればよく、実際的に表示手段上にリストアップする必要はない。したがって時間軸及びラグ軸も理論上の軸であればよい。ここで「類似線分」の概念は、本願明細書において定義するものである。類似線分は、予め定めた時間長さ以上類似度が予め定めた閾値以上あるときに、閾値以上ある類似度の部分の長さに対応する時間長さを有する線分として定義される。閾値を適宜に変更または調整することにより、ノイズを除去することが可能になる。なお閾値を設けたことによりノイズは除去できるものの、本来現れるべき類似線分が現れなくなる場合もある。しかしそのような場合であっても、今回の特徴量と過去

のすべての特徴量との間の類似度についての類似線分をリストアップするため、後に他の類似線分との関係から、本来現れるべき類似線分がないことを探索することができるので、リストアップの精度が下がることはない。

統合繰り返し区間決定ステップでは、リストアップされた複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある1以上の繰り返し区間を統合して一つの統合繰り返し区間を決定する。統合繰り返し区間決定ステップでは、時間軸の共通区間に存在するリストアップした類似線分どうしをそれぞれグルーピングにより統合して統合繰り返し区間と定める。そして複数の統合繰り返し区間を、共通区間の長さでグルーピングされる類似線分のラグ軸で見た位置関係とに基づいて複数種類の統合繰り返し区間列に分類する。より具体的には、リストアップされた複数の繰り返し区間の相互関係は、時間軸上の共通区間に対応する過去のラグ位置に1以上の繰り返し区間（類似線分）が存在するか否かと、そのラグ位置に対応する過去の時間帯において繰り返し区間（類似線分）が存在するか否かの関係である。これらの関係に基づいて、このステップでは、共通区間に対応する過去のラグ位置に1以上の繰り返し区間（類似線分）がある場合に、それらをその共通区間に繰り返し区間（類似線分）があるものと決定して、その繰り返し区間を統合繰り返し区間とする。その上で、統合繰り返し区間決定ステップでは、決定した複数の統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する。この分類化は、共通区間の長さの共通性と、共通区間に存在する繰り返し区間（類似線分）の位置関係と数との関係に基づいて行われる。この分類化により、種類の異なる繰り返し区間の構造化が実現できる。

なお統合繰り返し区間を用いると、類似度を求めた2番目以降の繰り返し区間に対応する統合繰り返し区間は求まるものの、最初の繰り返し区間が統合繰り返し区間列には含まれないことになる。そこで統合繰り返し区間決定ステップでは、統合繰り返し区間に含まれない最初の繰り返し区間を補足して統合繰り返し区間列を作成するようにしてもよい。

そしてサビ区間決定ステップで、複数種類の統合繰り返し区間列からサビ区間を決定する。このサビ区間決定ステップでは、例えば、統合繰り返し区間列に含まれる統合繰り返し区間の類似度の平均と、数と長さとに基づいて該統合繰り返



し区間列に含まれる統合繰り返し区間のサビらしさを求める。そして、最もサビらしさの高い統合繰り返し区間列に含まれる統合繰り返し区間をサビ区間として決定する。なおサビらしさの定め方は、一つに限定されるものではなく、よりよいサビらしさの基準に基づいて判断すれば、それだけ検出精度が高まるものは勿論である。

なお楽曲が転調を含んでいる場合には、次のようにする。まず特徴量抽出ステップでは、12次元クロマベクトルからなる音響特徴量を1転調幅ずつ11転調幅までシフトして得た転調幅の異なる12種類の音響特徴量を求める。次に類似度演算ステップでは、今回求めた音響特徴量と先に求めた全ての12種類の音響特徴量との間の類似度を、時刻 $t$ の今回の音響特徴量を表す12次元クロマベクトルとそれよりラグ $l$  ( $0 \leq l \leq t$ ) だけ過去の全ての12種類の音響特徴量を表す12次元クロマベクトルとの間の類似度として演算する。そして繰り返し区間リストアップステップでは、12種類の音響特徴量ごとに、一方の軸を時間軸 $t$ とし他方の軸をラグ $l$ とし、予め定めた時間長さ以上類似度が予め定めた閾値以上ある場合には類似度が予め定めた閾値以上である部分の長さに対応する時間長さを有する類似線分を時間軸を基準にした繰り返し区間としてそれぞれ12種類のリストをリストアップする。

統合繰り返し区間決定ステップでは、12種類のリストごとに、時間軸の共通区間に存在するリストアップした類似線分どうしをそれぞれグルーピングにより統合して統合繰り返し区間と定める。さらに12種類のリストについて定めた複数の統合繰り返し区間を共通区間の時間軸上の存在位置及び長さ、グルーピングされる類似線分のラグ軸で見た位置関係とに基づいて複数種類の転調を考慮した複数種類の統合繰り返し区間列に分類化する。このようにすると、転調を含んだ音楽音響データであっても、転調した部分の特徴量を11段階の転調幅のシフトでずらして類似度を求めるため、転調した部分の特徴量を正しく抽出することができる。その結果、繰り返し区間が転調されている場合でも、同じ特徴(Aメロ、Bメロ、サビ)の繰り返し区間であるか否かの判定を高い精度で行うことが可能になる。

ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響デー

タ中からサビ区間に対応する部分を検出して表示手段に表示する本発明のサビ区間検出装置は、音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出手段と、音楽音響データについて求めた複数の音響特徴量の相互間の類似度を求める類似度演算手段と、類似度に基づいて音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップ手段と、リストアップされた複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある1以上の繰り返し区間を統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定手段と、複数種類の統合繰り返し区間列からサビ区間を決定するサビ区間決定手段とを具備する。サビ区間を含む統合繰り返し区間列または複数種類の統合繰り返し区間列は、表示手段に表示される。そしてサビ区間を含む統合繰り返し区間列が他の統合繰り返し区間列とは異なる表示態様で表示される。このようにすると検出したサビ区間を他の繰り返し区間とは区別して明瞭に表示することができる。

なお本発明は、統合繰り返し区間列を表示手段に表示せずに、音響の再生手段でサビ区間を含む統合繰り返し区間列またはその他の統合繰り返し区間列を選択的に再生するようにしてもよいのは勿論である。

ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出する方法をコンピュータを用いて実現するために用いられるプログラムは、音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出ステップと、音楽音響データについて求めた複数の音響特徴量の相互間の類似度を求める類似度演算ステップと、類似度に基づいて音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップステップと、リストアップされた前記複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある1以上の繰り返し区間を統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定ステップと、複数種類の統合繰り返し区間列からサビ区間を決定するサビ区間決定ステップとを前記コンピュータに実行させるように構成されている。

## 図面の簡単な説明

図 1 は、本発明の楽曲再生方法の実施の形態の概念を示すブロック図である。

図 2 (A) は手動サビ区間ラベリング用エディタの画面の表示であり、図 2 (B) 及び (C) は試聴者行動分析手法で得られたヒストグラムの例を示す図である。

図 3 は、本発明の楽曲再生装置の一実施の形態の構成を示すブロック図である。

図 4 (A) はそれぞれ音楽音響データから自動的に楽曲中の複数の楽曲構造特徴区間を抽出して楽曲構造特徴区間指定データを作成する特徴区間指定データ生成手段を備えた本発明の楽曲再生装置の変形例を示すブロック図であり、図 4 (B) は他の変形例を示すブロック図である。

図 5 (A) 及び (B) は、インターフェースの表示画面の一例を示す図である。

図 6 は、インターフェースを実現する場合に用いるプログラムのアルゴリズムを示すフローチャートである。

図 7 は、通常の再生動作を行う際のインターフェースの動作のアルゴリズムを示すフローチャートである。

図 8 (A) ～ (C) は、サビ区間の選択を説明するために用いる図である。

図 9 (A) ～ (C) は、サビ区間の選択を説明するために用いる図である。

図 10 (A) ～ (C) は、繰り返し区間の選択を説明するために用いる図である。

図 11 (A) 及び (B) は、繰り返し区間の選択を説明するために用いる図である。

図 12 (A) ～ (C) は、繰り返し区間の選択を説明するために用いる図である。

図 13 は、繰り返し区間の選択を説明するために用いる図である。

図 14 は、実際に作成したインターフェースの画面を示す図である。

図 15 は、本発明のサビ区間検出方法で、転調を伴う楽曲中のサビ区間を検出

する一実施の形態の方法の処理ステップを示すフローチャートである。

図 16 は、本発明のサビ区間を検出する装置の実施の形態の一例の構成の概略を示すブロック図である。

図 17 は、図 16 の装置をコンピュータを利用して実現する場合に用いるプログラムのアルゴリズムの一例を示すフローチャートである。

図 18 は、螺旋状の音高知覚を説明するための図である。

図 19 は、1 2 次元クロマベクトルを説明するために用いる図である。

図 20 は、類似度の演算の考え方を説明するために用いる図である。

図 21 は、類似度の演算の考え方を説明するために用いる図である。

図 22 は、ある楽曲に対する類似線分、類似度  $r(t, 1)$ 、パラメータ空間  $R_{all}(t, 1)$  の概念図である。

図 23 は、実際に得られる類似線分の一例を示す図である。

図 24 は、類似線分の考え方を説明するために用いる図である。

図 25 は、類似線分の考え方を説明するために用いる図である。

図 26 は、類似線分の考え方を説明するために用いる図である。

図 27 は、類似線分の考え方を説明するために用いる図である。

図 28 は、類似線分を求める際の閾値の定め方を説明するために用いる図である。

図 29 は、類似線分を求める際の閾値の定め方を説明するために用いる図である。

図 30 は、類似線分の抽出方法を説明するために用いる図である。

図 31 は、繰り返し区間の統合化を説明するために用いる図である。

図 32 は、繰り返し区間の統合化を説明するために用いる図である。

図 33 は、繰り返し区間の統合化の例を示す図である。

図 34 は、繰り返し区間の統合化の例を示す図である。

図 35 は、統合繰り返し区間列の表示例を示す図である。

図 36 は、あるサビの転調前後での 1 2 次元クロマベクトルの違いを示す図である。

図 37 は、転調に対処するためのシフト処理を説明するために用いる図である。

図38は、転調処理のために12種類のリストを作成することを示す図である。

図39は、サビ区間の選定の仮定の一例を説明するために用いる図である。

図40は、サビ区間の選定の仮定の一例を説明するために用いる図である。

図41は、RWC-MDB-P-2001, No. 18の楽曲終了時点での正しいサビ検出結果を示す図である。

#### 発明を実施するための最良の形態

以下、本発明の実施の形態について詳細に説明する。まず、サビ区間を検出する場合の問題点について説明する。サビ区間の検出のためには、楽曲一曲分の音響信号データ中に含まれるすべてのサビ区間の開始点と終了点を求める必要がある。サビは、コーラス (chorus) あるいはリフレイン (refrain) とも呼ばれる。サビは、楽曲構造上、主題 (theme) を提示している部分を指す。そしてサビは、ときには伴奏の変化やメロディーの変形を伴いながら、通常は、楽曲中で最も多く繰り返される。例えば、典型的なポピュラー音楽の楽曲構造は、

{イントロ、サビ}

( (→第1の序奏部分 (Aメロ) [→第2の序奏部分 (Bメロ)] ) × n1 → サビ) × n2

[→間奏] [→第1の序奏部分 (Aメロ)] [→第2の序奏部分 (Bメロ)] → サビ × n3

[→間奏 → サビ × n4] [→エンディング]

のようになっている。このようにサビは、他のメロディよりも繰り返し回数が多くなっている。ここで、{a, b} はaかbかのいずれか一方、[a] はaが省略可能であることを表す記号である。そしてn1, n2, n3, n4 は繰り返し回数を表す正の整数である (多くの場合、 $1 \leq n1 \leq 2$ ,  $1 \leq n2 \leq 4$ ,  $n3 \geq 0$ ,  $n4 \geq 0$ )。イントロ (introduction) は前奏部分、Aメロ、Bメロ (verse A, verse B) は序奏部分を指す。

楽曲中で通常、最も多く繰り返されるサビの区間を検出するには、基本的には、ある楽曲中に含まれる複数の区間の繰り返し（繰り返し区間）を見つけ出し、最も出現頻度の大きい区間をサビ区間とすればよい。しかし、「繰り返し区間」とは言っても音響信号が完全に一致する状態で区間が繰り返される場合は希である。そのため、人間にとっては容易に繰り返しと分かる場合でも、計算機にとってはその判断が難しい。その際の主要な課題は、以下のようにまとめられる。

#### 課題 1：特徴量と類似度の検討

ある区間の音響信号とその区間の繰り返し区間と考えられる他の区間の音響信号とが完全に一致しない場合には、ある区間が繰り返されているということを判断するために、各区間から求めた特徴量相互間の類似度を判断しなければならない。その際、繰り返しがあると判断できるためには、繰り返す度にその区間内の音響信号の細部が多少異なっても（メロディーが変形したり、伴奏のベース、ドラム等が演奏されなくなったりしても）、各区間の特徴量間の類似度は高い必要がある。しかしながら、各区間のパワースペクトルを直接特徴量とした場合には、この類似度の判断が困難である。

#### 課題 2：繰り返しの判断基準

類似度がどれくらい高ければ繰り返しとみなせるかという基準は、楽曲に依存して変わる。例えば、似た伴奏が多用される楽曲では、全体的に多くの部分の類似度が高くなる。そのため、比較する各区間の類似度がかなり高い類似度でなければ、それらの区間がサビに関連する繰り返し区間であると判断しない方がよい。逆に、サビが繰り返されるときに、伴奏が大きく変化するような楽曲では、比較する各区間の類似度がやや低くても繰り返し区間であると判断する方がよい。こうした基準を、ある楽曲に特化して人間が手作業で設定するのは容易である。しかしながら、幅広い楽曲からサビ区間を自動的に検出するためには、サビ区間の検出の基準を、現在処理中の楽曲に応じて自動的に変える必要がある。このことは、あるサビ区間の検出方法の性能を評価する場合に、その方法で数曲のサンプル曲についてサビ区間の検出ができたからといって、必ずしもその方法に汎用性があるとは限らないということを意味する。

#### 課題 3：繰り返し区間の端点（開始点と終了点）の推定

サビ区間の長さ（区間長）は楽曲ごとに異なるため、各区間長と共に、どこからどこまでがサビであるかを推定しなければならない。その際、サビの前後の区間も一緒に繰り返すことがあるため、端点の推定は、楽曲中の様々な箇所の情報を統合して行う必要がある。例えば、（A B C B C C）のような構造の楽曲の場合（A, B, CはそれぞれAメロ、Bメロ、サビの区間とする）、単純に繰り返し区間を探すと、（B C）が一つのまとまった区間として見つかる。この場合、最後のCの繰り返し情報に基づいて、（B C）の内のCの区間の端点を推定する、といった処理が求められる。

#### 課題4：転調を伴う繰り返しの検出

転調後の区間は、一般に特徴量が大きく変わるために、転調前の区間との類似度が低くなり、繰り返し区間と判断するのが困難となる。特に、転調は曲の後半のサビの繰り返しで起きることが多く、そうした繰り返しを的確に判断することは、サビの検出において重要な課題である。

本発明では、以上の課題を解決しつつ、基本的に楽曲中で多く繰り返される区間をサビとして検出する。以下の実施の形態の説明においては、入力として、音楽のモノラルの音響信号を対象とし、混合音中の楽器の数や種類には特に制限を設けない。ステレオ信号の場合には、左右を混合してモノラル信号に変換するものとする。以下の実施の形態では、以下のことを仮定する。

仮定1：演奏のテンポは一定でなく変化してもよい。しかしサビの区間は、毎回ほぼ類似したテンポで、一定の長さの区間として繰り返し演奏される。その区間は長い方が望ましいが、区間長には、許容される適切な範囲（現在の実装では、7.7～40sec）がある。

仮定2：前述した楽曲構造の例の、

$((\rightarrow A \text{メロ} [\rightarrow B \text{メロ}]) \times n1 \rightarrow \text{サビ}) \times n2$

に相当するような、長い繰り返しがある場合、その末尾の部分がサビである可能性が高い（図39参照）。

仮定3：サビ区間内では、その区間の半分程度の長さの短い区間が繰り返されることが多い。そのため、ある繰り返し区間内にさらに区間の短い繰り返し区間がある場合には、その区間がサビである可能性が高い（図40参照）。以上は、

多くのポピュラー音楽に当てはまる妥当な仮定である。本実施の形態においては、上記課題と仮定を前提にしている。なお以下の説明では、先に本発明の楽曲再生方法の実施の形態を説明し、次に本発明のサビ区間の検出方法の実施の形態について説明する。

図1は本発明の楽曲再生方法の実施の形態の概念を示すブロック図である。これらの実施の形態では、楽曲構造上意味を持つ区間（サビ区間及び繰り返し区間）の先頭に、音楽の再生位置を自動ジャンプさせる。基本的には、音楽音響信号や標準MIDIファイル等の音楽音響データの楽曲構造を事前に解析しておき、その中で試聴者（ユーザ）が関心を持つ部分（サビの部分、繰り返し演奏される部分）すなわち楽曲構造特徴区間へ演奏箇所を自動的にジャンプできる機能を提供する。音楽音響データは、ローカルな電子計算機（サーバ）等やCDプレーヤ等の音楽再生器上に用意されていてもよく、またネットワークからダウンロードされるものでもよい。

#### [第1のステップの説明]

具体的には、再生の対象となる音楽音響データに関して、楽曲中の楽曲構造特徴区間を指定するための複数の楽曲構造特徴区間指定データを予め用意する（第1のステップ）。最も単純には、音楽音響データを再生しながら複数の楽曲構造特徴区間を人が特定し、その特定結果を基準にして複数の楽曲構造特徴区間を指定するための複数の楽曲構造特徴区間指定データを作成する。この作業はマニュアルでの作業になるため、時間がかかるものの、実際に人間（例えば、一般人だけでなく、音楽を深く理解している専門家や音楽製作者）が判断して楽曲構造特徴区間を特定するため、例えばサビ区間が分かり難い楽曲でも必要な楽曲構造特徴区間指定データを得ることができる。なお本実施の形態では、複数の楽曲構造特徴区間のそれぞれの始まり位置と終了位置を、実際に演奏を開始してからその位置に達するまでに要する時間で表した位置データとして表現している。そして本実施の形態では、複数の楽曲構造特徴区間指定データを音楽音響データと組合せて一つのセットにしている。

図1の実施の形態では、楽曲構造特徴区間を、自動サビ区間検出手法、手動サビ区間ラベリング手法、試聴者行動分析手法のいずれか一つの方法を用いて楽曲



構造特徴区間を特定する。以下各手法を説明する。

〔自動サビ区間検出手法〕

この手法では、楽曲中のサビ区間や繰り返し区間の開始点と終了点の一覧を自動検出する。本手法は、人手がかからないため多数の楽曲に適用可能である。自動検出では検出結果に誤りが含まれることもあるが、実用上は、完全な精度でなくても試聴者が再生位置を見つける手掛かりとなるため、従来の試聴機よりは便利となる。このための手法の詳細については、後に詳しく説明する。

〔手動サビ区間ラベリング手法〕

サビや楽曲構造をラベリング可能なエディタ上で、人間が音楽を聴きながら手作業で指定する。精度は高いが、労力と時間がかかる。この手動ラベリングは、自動検出が困難な楽曲、音楽ジャンルに対しても有効である。なお、自動サビ区間検出の結果を利用して、それをエディタ上で手作業で修正することもできる。このためのエディタとして、発明者は楽曲を分割して各部にサビ、第1のメロディ（Aメロ）、第2のメロディ（Bメロ）、イントロ、間奏、エンディング等をラベリングできる、楽曲構造ラベリング用エディタを開発した。エディタの画面を図2（A）に示す。この図の上半分に示された横軸は時間軸（sec）で楽曲全体を表示している。一番上のミニウインドウがパワー変化、二番目のミニウインドウが楽曲各部のラベリング結果、三番目のミニウインドウの最上段がサビ区間の一覧であり、下5段が繰り返し構造を表す。下半分は、楽曲各部のラベリング結果を、楽曲全体の時間軸を折り返ししながら拡大表示している。このエディタ上では、ユーザは縦棒表示で表されるカーソルの位置の情報を編集する。楽曲の再生・停止は自由に制御することができ、カーソルは楽曲の再生位置を指す。ユーザはカーソルを自由に移動しながら、カーソル前後の音を繰り返し聞いたりして、適切な箇所に、楽曲の各部の区切りを入れることができる。そして、区切りの間の区間をラベリングする。なお、このエディタは、事前にビートや小節の区切りをラベリングしておき、エディタ上ではビートや小節単位でカーソルを移動することで、ラベリング作業を迅速に行う機能も持つなど、効率のよい作業環境を提供している。このエディタ上での編集結果として、楽曲中のサビ区間や繰り返し区間の開始点と終了点の一覧を得ることができる。

### 〔試聴者行動分析手法〕

この手法では、過去の多数の試聴者が聴いた箇所を記録しておき、試聴者が多く聴いた区間を分析する。そうして求めた区間を、サビ区間（試聴者が関心を持つ区間）であると仮定して推定する。ポピュラー音楽以外でも、多く聴いた区間は典型的な試聴者が聴きたい区間となっていることが多く、幅広い音楽ジャンルに有効である。本手法は、繰り返し利用されるまで、実質的に従来の試聴機とあまり変わらない機能しか提供できないが、事前に楽曲内容の分析やラベリングを必要としない利点を持つ。なお、上記二つの手法と組み合わせることも可能である。具体的には、以下のように行う。基本的には、楽曲中の各箇所の再生回数のヒストグラムを作成し、ヒストグラムから多く再生されている箇所を求め、そこをサビ区間であると推定する。ヒストグラムの例を図2（B）及び（C）に示す。まず、最初はまったく記述がない状態で始め、従来のように典型的な再生操作ボタンを使って（早送りボタンを何度も押して途中を飛ばしながら）、試聴者はサビ区間等を手探りで見つけながら聴く。その際に、楽曲全体を非常に短い区間（例えば10 msec）に区切り、その各区間ごとの再生回数をカウントしておく（ヒストグラムを作成する）。通常、サビの先頭が見つかる、サビ全体が再生し終わるまで試聴者は聞くため、これを多数の試聴者が行くと、多くの試聴者が興味を持って再生した区間が、ヒストグラム上で長い区間に渡って高頻度な区間（多く再生された区間）として求めることができる。さらに拡張として、再生回数をカウントする際に、ボタン操作をせずに連続して再生した区間の長さに応じた重みを付け、長く再生するほど、大きな値がヒストグラムに加算されるようにする。これにより、サビの探索中は短い区間しか再生されないために小さい値がヒストグラムに加算され、サビが見つかる、長い区間を再生されるために大きい値がヒストグラムに加算される。こうして、ヒストグラム中のサビの区間の値をより高くすることができ、より適切にサビの区間が求まる。

なお上記の処理は、単独の試聴機で実行するだけでなく、試聴機同士をインターネット等の広域ネットワークで接続し、同一楽曲に対して、世界中の試聴者（ユーザ）がどのように操作して聴いたのかのヒストグラムを集計することもできる。これにより、より多くの試聴者の行動分析結果がヒストグラムに反映され、

サビの区間を求める精度が向上する。なお試聴者行動分析手法の場合には、ヒストグラムを視覚化するのが効果的であり、過去の試聴者がどこをよく聴いたのかを視覚的に見ながら操作できる。

試聴者行動分析手法の結果を使うには、図2(B)に示すようにヒストグラムのある閾値以上をサビの区間(典型的な試聴者が関心を持つ区間)として、ヒストグラムの上か下に、明示的に表示すると便利である。このようにして定めた区間を楽曲構造特徴区間として本発明を適用すれば、ユーザは他の手法で求めた楽曲構造特徴区間と同様に、インターフェースによる選択に従って再生位置をジャンプさせることが可能である。あるいは図2(C)に示すように、このような区間に代わる別の表示方法として、ヒストグラムを平滑化してスムーズにした関数を求め、それを時間軸方向に一次微分した関数(導関数)が、ある閾値以上になっている時刻に線分(複数存在する)を表示してもよい。それらの線分は、ヒストグラムが急激に上昇した箇所を表しており、多くの試聴者が聴き始める箇所(典型的な試聴者が関心を持つ区間の開始時刻)を表している。この場合でも、ユーザがインターフェースにより線分を選択することにより、それらの線分の時刻に対応する再生位置をジャンプして再生することが可能である。

#### [第2のステップ及び第3のステップの説明]

次に本実施の形態の方法では、複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するインターフェースを用意する(第2のステップ)。このインターフェースについては、後に詳しく説明する。そして音楽音響データ再生装置を用いて音楽音響データの再生を開始した後、操作者がインターフェースを用いて任意の楽曲構造特徴区間を選択すると、選択した任意の楽曲構造特徴区間に対応する楽曲構造特徴区間指定データに基づいて、音楽音響データ再生装置による音楽音響データの再生位置を任意の楽曲構造特徴区間に変更する(第3のステップ)。この再生位置の変更がいわゆる楽曲構造特徴区間の先頭へのジャンプである。図1の実施の形態では、インターフェースに画像表示部を設けて、楽曲の内容を反映しながら楽曲構造特徴区間を視覚により見えるようにしている。なおこの視覚化も後に詳しく説明する。この実施の形態によれば、具体的には、「サビ区間の頭出し」、「前の楽曲構造の区間(繰り返し区間)の頭出し」、「次の楽曲構

造の区間（繰り返し区間）の頭出し」が可能となり、試聴者がサビの部分だけ聴いたり、前後の楽曲構造の先頭へとジャンプして聴いたりすることができるようになる。

図 3 は本発明の方法を実施する本発明の楽曲再生装置の一実施の形態の構成を示すブロック図である。この楽曲再生装置 1 は、音楽音響データを再生する音楽音響データ再生装置 3 を備えて楽曲中の複数の楽曲構造特徴区間から選択した任意の楽曲構造特徴区間を再生する機能を備えている。この例では、例えばネットワークを介して送られてきた楽曲構造特徴区間指定データ付の音楽音響データがデータ分離手段 5 に入力される。データ分離手段 5 は、楽曲構造特徴区間指定データ付の音楽音響データから音楽音響データと楽曲構造特徴区間指定データとを分離し、音楽音響データについては音楽音響データ格納手段 7 に格納し、楽曲構造特徴区間指定データについては楽曲構造特徴区間指定データ格納手段 9 に格納する。制御部 11 は、音楽音響データ再生装置 3 とインターフェース 13 とを内蔵してディスプレイ 15 を備えた画像表示部 17 をコントロールする。制御部 11 の内部には、指定データ付与手段 12 が含まれている。音楽音響データ再生装置 3、制御部 11、指定データ付与手段 12、インターフェース 13 及び画像表示部 17 の主要部はプログラムによって動作可能になる図示しないコンピュータにより実現される。

楽曲構造特徴区間指定データ格納手段 9 は、再生する音楽音響データに対応して予め定められた複数の楽曲構造特徴区間を指定するための楽曲構造特徴区間指定データを格納する。そしてインターフェース 13 は、後に詳しく説明するように、複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択する機能を有する。また指定データ付与手段 12 は、インターフェース 13 によって選択された任意の楽曲構造特徴区間を指定する楽曲構造特徴区間指定データを楽曲構造特徴区間指定データ格納手段 9 から読み出して音楽音響データ再生装置 3 に与える。音楽音響データ再生装置 3 は、インターフェース 13 からの指令に従って動作し、特に指定データ付与手段 12 から楽曲構造特徴区間指定データが与えられると、指定データによって指定された楽曲構造特徴区間に再生位置を変更する。この楽曲再生装置 1 では、楽曲の再生中であっても、インターフェース 13 によって

選択された任意の楽曲構造特徴区間を再生することができる。なお音楽音響データ再生装置 3 は、楽曲の再生中でなくても、インターフェースによって選択された任意の楽曲構造特徴区間を再生できるように構成されていてもよいのは勿論である。

図 4 (A) は、音楽音響データから自動的に楽曲中の複数の楽曲構造特徴区間を抽出して楽曲構造特徴区間指定データを作成する特徴区間指定データ生成手段 2 1 を備え、またインターフェース 1 3 が独立して存在している本発明の楽曲再生装置 1 の変形例を示している。このような特徴区間指定データ生成手段 2 1 を備えていれば、既存の楽曲を再生する場合にも本発明を適用することが容易になる。なお特徴区間指定データ生成手段 2 1 は、前述の自動サビ区間検出手法を実施できるように構成すればよい。また図 4 (A) の実施の形態においては、1 以上の選択ボタンを備えたインターフェース 1 3 が独立した装置として用意されている。このインターフェース 1 3 では、選択ボタンがいわゆる指で物理的に操作するマニュアルスイッチにより構成されている。なお選択ボタンは、タッチパネル上に画面表示されているものを指やペンでクリックして操作したり、マウスでクリックして操作するタイプのものでもよく、その構造は任意である。ペンで操作するタイプのものでもよく、その構造は任意である。

また図 4 (B) に示すように、インターフェース 1 3 を含んだ画像表示部 1 7 を独立した装置として構成してもよいのは勿論である。この例では、音楽音響データと楽曲構造特徴区間指定データのそれぞれを別個にダウンロードしてそのまま音楽音響データ格納手段 7 と楽曲構造特徴区間指定データ格納手段 9 に格納するようにしている。

前述の図 3 の実施の形態では、いずれもディスプレイ 1 5 上にインターフェース 1 3 の操作部及び表示部が表示される構成を有している。インターフェースの動作を制御する手段は、画像表示部 1 7 内に内蔵されている。図 5 (A) は、ディスプレイ 1 5 上に表示したインターフェース 1 3 の操作部及び表示部の構成の一例を示す図である。この例では、上側のウィンドウ W 1 内に、音楽音響データの長さを時間軸を用いた画像表示 A により表示し、併せて音楽音響データ再生装置 3 が音楽音響データを再生している音楽音響データの再生位置と複数の楽曲構

造特徴区間とを時間軸を用いた画像表示Bにより表示し、下側のウインドウW2内に各種の選択ボタンSW1～SW10もディスプレイ15上に画像表示Cによって画像表示部17が表示している。

下側のウインドウW2には、左側から順に、停止選択ボタンSW1、一時停止選択ボタンSW2、再生選択ボタンSW3、前の曲の頭出し選択ボタンSW4、早戻し選択ボタンSW5、早送り選択ボタンSW6、次の曲の頭出し選択ボタンSW7が並んでおり、これらのボタンの上には慣例的な記号が描かれている。これらの選択ボタンは、既存のインターフェースで用いられているものと同じである。停止選択ボタンSW1の上側には、楽曲の先頭からの経過時間が表示されている。

本実施の形態で用いるインターフェースでは、ウインドウW2内にさらに自動ジャンプボタンと呼ばれる3つの選択ボタンSW8～SW10を備えている。左から順に、「前の楽曲構造の区間の頭出し」用の第1の繰り返し区間選択ボタンSW8、「次の楽曲構造の区間の頭出し」用の第2の繰り返し区間選択ボタンSW9及び「サビ区間の頭出し」用のサビ区間選択ボタンSW10である。

サビ区間選択ボタンSW10を押すと、現在の再生位置より後方の（後方になれば最初の）サビ区間を探索し、その開始点にジャンプする。通常、サビは楽曲中で複数回繰り返されるが、このボタンSW10を押す度にそれらの間を順にジャンプできる。他の二つのボタンSW8及びSW9を押すと、現在の再生位置の直後もしくは直前に位置する繰り返し区間の開始点を探索し、その先頭にジャンプする。探索時には、区間の終了点は無視する。

次に、上側にウインドウW1（楽曲構造表示ウインドウ）の構成について説明する。最上段にサビ区間の表示バーB1があり、その下に5段の繰り返し区間表示バーB2～B6が視覚により認識できるように表示されている。この例では、5段の繰り返し区間表示バーを用いているが、表示バーの数は任意であり、この実施の形態に限定されるものではない。サビ区間の表示バーB1の区間を表示する色と同じ色のマークがサビ区間選択ボタンSW10の上に表示しており、また区間表示バーB2～B6の区間を表示する色と同じ色のマークが二つのボタンSW8及びSW9の上に表示されている。これにより操作者がボタンを選択する際

の誤認動作が防止できる。この例は、5種類の区間が繰り返し演奏されていることを表示している。最下段の細い横棒の表示Aは、楽曲中の経過時間を知らせる再生位置スライダである。またこの例では、6本の表示バーB 1乃至B 6を横切るように縦型表示バーIを設けてある。この縦型表示バーIは、再生位置スライダAの表示マークaと同期して変位して、現在の再生位置を表示する。このような表示を用いる場合、各区間を直接クリック（タッチパネル使用時にはタッチ）して音楽音響データを再生したり、再生位置スライダをクリックして再生位置の変更をしたりすることが可能である。以上のインターフェースと機能により、試聴者は、イントロを少し聴いた後に、ボタンを押すだけでサビを試聴することが可能となる。なお、ウインドウW 2に「次のサビ区間の頭出し」用の第1のサビ区間選択用ボタンと「前のサビ区間の頭出し」用の第2のサビ区間選択用ボタンの二つを用意してもよい。この実施の形態では、以下の理由から一つのボタンとしている。（1）現状の「サビ区間の頭出し」用のサビ区間選択ボタンSW 10を連打すれば、すべてのサビ区間を一巡した後にまた最初の区間に戻るなので、短時間で望みの箇所へ移動できる。（2）瞬時に過去のサビ区間へ戻る必要がある場合には、楽曲構造表示ウインドウW 1で区間を直接クリックする手段が用意されている。また、楽曲全体の構造を視覚的に把握しながら、様々な箇所を選択的に試聴できる。

この例のように、サビ区間や楽曲中の繰り返し区間の構造を視覚化する（いずれか一方でも良い）と、この視覚化された画面から、試聴者は、イントロ、第1の種類のメロディ（Aメロ）、第2の種類のメロディ（Bメロ）、サビ、間奏等の相互の位置関係が把握できることが多い。

図5（B）は他のインターフェースの表示態様を示している。図5（B）のインターフェースでは、サビ区間選択ボタンが次のサビ区間にジャンプする場合に用いる第1のサビ区間選択ボタンSW 10の他に、前のサビ区間にジャンプさせる第2のサビ区間選択ボタンSW 11を備えている。その他の点は、図5（A）のインターフェースと同じである。

図6は、上記実施の形態で用いるインターフェース13を実現する場合に用いるプログラムのアルゴリズムを示すブロック図である。なお図6のアルゴリズム

では図5（B）の例に対応している。すなわち、サビ区間選択ボタンが次のサビ区間にジャンプする場合に用いる第1のサビ区間選択ボタンとは別に、前のサビ区間にジャンプさせる第2のサビ区間選択ボタンを備えているものと仮定している。

まずステップS T 1では、サビ区間及び繰り返し区間を表示バーB 1乃至B 6として視覚化する。次にステップS T 2で再生位置を縦表示バーIと表示マークaによって表示する。これらの表示が終了すると、ステップS T 3において、停止選択ボタンS W 1、一時停止選択ボタンS W 2及び再選選択ボタンS W 3のいずれが押されているか（クリックされているか又は選択されているか）の判定が行われる。そして停止選択ボタンS W 1が押されていれば、音楽音響データの再生位置をデータの先頭に設定して、再生状態を停止に設定する。また一時停止選択ボタンS W 2が押されていれば再生を一時停止に設定する。更に再生選択ボタンS W 3が押されていれば、再生状態を再生に設定する（なおこのときの再生速度は普通である。）。ステップS T 4が終了すると、ステップS T 2へと戻り、縦表示バーIと表示マークaの位置を再生位置に合わせて変更する。いずれの選択ボタンも押されていない場合には、ステップS T 5へと進む。ステップS T 5では、早送り選択ボタンS W 6または早戻し選択ボタンS W 5が押されているか否かの判定が行われる。早送り選択ボタンS W 6、早戻し選択ボタンS W 5のいずれかが押されている場合には、ステップS T 6へと進んでボタンが押されている間だけ再生速度を早送りまたは早戻しに変更し、次にステップS T 7で再生状態を再生に設定する。その結果、再生をしながら早送りまたは早戻しが実行される。なお再生速度の変更に応じた再生位置の変更はステップS T 2で表示され、縦表示バーIと表示マークaの移動速度も変更される。

早送り選択ボタンS W 6、早戻し選択ボタンS W 5のいずれもが押されていない場合には、ステップS T 8へと進む。ステップS T 8では、次のサビ区間の頭出しのために第1及び第2のサビ区間選択ボタンS W 1 0またはS W 1 1が押されたか否かが判断される。第1及び第2のサビ区間選択ボタンS W 1 0またはS W 1 1のいずれかが押されていれば、ステップS T 9へと進む。ステップS T 9では、再生位置または再生位置を次のサビ区間の先頭に変位させることを設定す



る。そしてステップS T 7を経由してステップS T 2へと戻り、縦表示バー I と表示マーク a のジャンプまたは変位が表示される。ステップS T 8で選択ボタン S W 1 0 または S W 1 1 が押されていないことが確認されると、ステップS T 1 0 へと進む。ステップS T 1 0 では、前の繰り返し区間の先頭または後の繰り返し区間の先頭にジャンプするための第 1 及び第 2 の繰り返し区間選択用ボタン S W 8 及び S W 9 が押されているかが判定される。もし第 1 及び第 2 の繰り返し区間選択用ボタン S W 8 及び S W 9 のいずれかが押されていれば、ステップS T 1 1 へと進む。ステップS T 1 1 では、現在の再生位置又は再生位置から次のまたは前の繰り返し区間の先頭に再生位置を変位させる。そしてステップS T 7 へと進んで再生状態を再生に設定し、ステップS T 2 で再生位置の表示が変更される。ステップS T 1 0 においては、第 1 及び第 2 の繰り返し区間選択用ボタン S W 8 及び S W 9 のいずれもが押されていない場合には、ステップS T 1 2 へと進む。

ステップS T 1 2 は、ディスプレイ 1 5 上の各区間表示がそのまま選択ボタンになるようにする場合に設けられる。この場合にも各区間が直接クリック（タッチ）された場合には、そのクリック（タッチ）された区間の先頭に再生位置を変位する。その後ステップS T 7 へと進んで再生状態を再生に設定し、ステップS T 2 へと進んで再生位置の変更が表示される。

ステップS T 1 2 で選択ボタンがクリックされていない場合には、ステップS T 1 4 へと進む。ステップS T 1 4 では、再生位置スライダのマーク a をクリック（タッチ）してマーク a を移動させる指示操作が行われたか否かが判断される。操作が行われると、ステップS T 1 5 へと進んでスライダのマーク a の移動先に再生位置を設定し、その後ステップS T 7 で再生状態を再生に設定してステップS T 2 へと戻る。

図 7 は、選択ボタン S W 1 ～ S W 7 を用いて操作する場合のアルゴリズムを示すフローチャートである。図 6 に示したアルゴリズムに従った動作と図 7 のアルゴリズムに従った動作とは並行して進行している。選択ボタン S W 8 ～ S W 1 1 が押されなければ、図 6 に従って再生が行われ、その再生の途中で選択ボタン S W 8 ～ S W 1 1 等が押されると、図 6 のアルゴリズムの所定のステップが実行さ

れる。

なお図 6 及び図 7 において、「再生状態」とは、停止、一時停止及び再生の状態を含み、「再生位置」とは楽曲のファイルの先頭からの経過時間であり、「再生速度」には通常の再生速度と、早送りの再生速度と、早戻しの再生速度が含まれる。

図 8 乃至図 13 を用いてインターフェース 13 の動作状態を説明する。なお表示バー B1～B6 において、黒色の部分と灰色の部分、サビ区間及び繰り返し区間の楽曲構造特徴区間を示している。またこれらの図において、押された状態にある選択ボタンには、理解を容易にするためハッチングを付してある。最初、図 8 (A)～(C) 及び図 9 (A)～(C) を用いて、サビ区間の頭出し動作を説明する。図 8 (A) に示す状態が再生選択ボタン SW3 を押して、ある音楽音響データの再生を開始した直後の状態である。この状態では、縦表示バー I 及び表示マーク a が、再生の最初の位置にある。図 8 (B) は再生状態を続けている状態を示している。そしてこの状態でサビ区間選択ボタン SW10 が 1 回押されると、図 8 (C) のようになる。この状態では、後方のサビ区間の表示の先頭に縦表示バー I 及び表示マーク a はジャンプし、再生位置も最初のサビ区間の先頭に変位する。しばらく再生状態を継続した後再度サビ区間選択ボタン SW10 を押すと、図 9 (A) に示すように、更に 2 番目のサビ区間の表示の先頭に縦表示バー I 及び表示マーク a は移動またはジャンプし、再生位置も 2 番目のサビ区間の先頭に変位する。そして更にサビ区間選択ボタン SW10 を押すと、図 9 (B) に示すように 3 番目のサビ区間の表示の先頭に縦表示バー I 及び表示マーク a は移動またはジャンプし、再生位置も 3 番目のサビ区間の先頭に変位する。さらにサビ区間選択ボタン SW10 を押すと、次にサビ区間が存在しないため、図 9 (C) に示すように、最初のサビ区間の先頭に縦表示バー I 及び表示マーク a は移動またはジャンプし、再生位置も最初のサビ区間の先頭に変位する。

次に図 10 及び図 11 を用いて繰り返し区間をジャンプする動作を説明する。図 10 (A) に示す状態が再生選択ボタン SW3 を押して、ある音楽音響データの再生を開始した直後の状態である。この状態では、縦表示バー I 及び表示マーク a が、再生の最初の位置にある。図 10 (B) は再生状態を続けている状態を示

している。そしてこの状態で第2の繰り返し区間選択ボタンSW9が1回押されると、図10(C)のようになる。この状態では、最も近い後方の繰り返し区間の〔表示バーB4の2番目の繰り返し区間〕の表示の先頭に縦表示バーI及び表示マークaはジャンプし、再生位置も最も近い後方の繰り返し区間〔表示バーB4の2番目の繰り返し区間〕の先頭に変位する。しばらく再生状態を継続した後再度第2の繰り返し区間選択ボタンSW9が1回押されると、図11(A)に示すように、この状態で最も近い後方の繰り返し区間の〔表示バーB5の1番目の繰り返し区間〕の表示の先頭に縦表示バーI及び表示マークaはジャンプし、再生位置も最も近い後方の繰り返し区間〔表示バーB5の1番目の繰り返し区間〕の先頭に変位する。そして更に第2の繰り返し区間選択ボタンSW9が1回押されると、図11(B)に示すように、この状態で最も近い後方の繰り返し区間の〔表示バーB5の2番目の繰り返し区間〕の表示の先頭に縦表示バーI及び表示マークaはジャンプし、再生位置も最も近い後方の繰り返し区間〔表示バーB5の2番目の繰り返し区間〕の先頭に変位する。

次に図12及び図13を用いて繰り返し区間を前方にジャンプする動作を説明する。図12(A)に示す状態が再生選択ボタンSW3を押して、ある音楽音響データの再生を開始した直後の状態である。この状態では、縦表示バーI及び表示マークaが、再生の最初の位置にある。図12(B)は再生状態を続けている状態を示している。そしてこの状態で第1の繰り返し区間選択ボタンSW8が1回押されると、図12(C)のようになる。この状態では、最も近い前方の繰り返し区間の〔表示バーB4の1番目の繰り返し区間〕の表示の先頭に縦表示バーI及び表示マークaはジャンプし、再生位置も最も近い前方の繰り返し区間〔表示バーB4の1番目の繰り返し区間〕の先頭に変位する。この状態で再度第1の繰り返し区間選択ボタンSW8が1回押されると、図13に示すように、この状態で最も近い後方の繰り返し区間の〔表示バーB6の1番目の繰り返し区間〕の表示の先頭に縦表示バーI及び表示マークaはジャンプし、再生位置も最も近い前方の繰り返し区間〔表示バーB6の1番目の繰り返し区間〕の先頭に変位する。

このようにして選択ボタンSW8～SW10を任意に選択することにより、インターフェース13によって選択された任意の楽曲構造特徴区間を再生すること

ができる。

図14に発明者が実際に作成したインターフェースの画面の一例を示す。

次に本発明をより具体的に実施した実機とその結果について説明する。

実機のインターフェースは、サビ区間検出手法を用いて得たサビ区間と繰り返し構造の記述を含むファイルを利用する形で動作する。

実機のGUI部分、音楽音響再生装置のエンジン部分、音響出力デバイス制御部分は、拡張性が高くなるように、分散環境で動作する別々のプロセスとして実装した。そのために、音響信号や各種制御情報をネットワーク上で効率よく共有することを可能にするネットワークプロトコルRACP (Remote Audio Control Protocol) を設計し、それに基づいて実装した。RACPは、RMCP (遠隔音楽制御用プロトコル) を音響信号の伝送用に拡張したプロトコルである。

実機を、二つの提案機能 (ジャンプのための選択ボタンSW8~SW10と楽曲構造表示) の有無に応じて、四つの条件で運用した。試聴対象には、「RWC 研究用音楽データベース: ポピュラー音楽」の100曲 (RWC-MDB-P-2001 No.1~100) の中から、試聴者が初めて聴く楽曲を選んだ。ただし、自動サビ区間検出手法によって正しく得られた楽曲構造特徴区間指定データを用いた (自動サビ区間検出手法による100曲中の正答曲数は、80曲であった)。参考までに、これら100曲に対して、開発した手動サビ区間ラベリング用エディタを用いて人間 (音楽大学卒業生) が手作業でラベリングするのに、事前にビートや小節の区切りをラベリングしてあるところから始めて、約100時間を要した。以下、四つの条件とその運用結果を述べる。ここでは、条件間の比較のために、視覚化された楽曲構造上の区間を直接クリックして再生する機能は用いていない。

[条件1] 提案機能が一切ない (従来に相当)。

試聴者は、曲の頭の部分がサビでない場合、少し聴いては早送り選択ボタンを押すという動作を、サビが出てくるまで5~10回程度繰り返した。少しずつ聴きながら早送りすることは、時間がかかって煩わしいものの、楽曲の雰囲気をつかみたいときには有効であった。

〔条件2〕 選択ボタンSW8～SW10はないが、楽曲構造表示はある。

どこまで再生位置を早送りすればよいかのフィードバックが得られる点が効果的であり、試聴者は条件1よりも便利だと評価した。ただし、画面表示によって現在の再生位置よりも先が見えているだけに、そこまで飛ばしたいという欲求が強くおきる傾向があった。

〔条件3〕 選択ボタンSW8～SW10はあるが、楽曲構造表示がない。

最初にイントロを聴いた後に、直接サビ区間選択ボタンSW10を押す聴き方と、「次の楽曲構造の区間の頭出し」に用いる第2の繰り返し区間選択ボタンSW9を押しては少し聴くという操作を繰り返し、サビが出てきたらそこをじっくり聴くという聴き方が主だった。効率よく飛ばしながら聴ける点が評価され、条件2より好まれていた。

〔条件4〕 提案機能がすべて有効である（実施の形態に相当）。

条件2、条件3の利点が共に得られる試聴方法であり、最も便利だと評価された。条件3のような聴き方に加え、楽曲構造上を自在に行き来しながら聴く傾向が強くなり、サビを聴いた後に、第1の種類のメロディ（Aメロ）に戻って聴いたり、楽曲の後半のサビの繰り返しへ飛んで聴いたりしていた。

条件3は、通常のメディアプレーヤに三つの選択ボタンSW8～SW10を追加した場合に相当し、楽曲構造表示がなくても通常のプレーヤより便利であることがわかった。また条件4から、視覚化はさらにその操作を助け、楽曲の様々な箇所を聴く上で有効であることがわかった。さらに、ボタンの機能や表示ウインドウの中身に関して一切説明を受けていない試聴者が、条件4で利用した場合でも、短時間の使用でこれらの意味を把握することが確認された。

以上から、インターフェースが実用的に機能し、試聴者は、楽曲構造表示の助けを得ながらジャンプ用の選択ボタンSW8～SW10を押して、インタラクティブに楽曲再生に介入できることを確認した。

上記実施の形態は、インターフェースの操作部をディスプレイ上に形成するものであるが、アナログスイッチ等の実際のスイッチを用いてインターフェースを構成してもよいのは勿論である。

また上記実施の形態は、いわゆるパーソナルコンピュータにプログラムをイン

ストールすることにより実現される。しかしながら例えば、いわゆる、ラジオカセット、CDプレーヤ、DVDプレーヤ等の各種の音楽音響再生器においても本発明を適用できるのは勿論である。

本実施の形態で用いるインターフェースは、一般的な音楽再生インターフェースとして有用であり、従来の楽曲単位での操作体系に対し、楽曲内部の区間単位での操作体系が追加されたものである。従来は楽曲単位で興味のない曲を飛ばせたが、楽曲内部の興味のない箇所を飛ばすのは容易でなかった。本発明のインターフェースにより、原曲の時系列に沿わずに、「好きなところを聴きたいように聴ける」ようになったメリットは大きい。なお、楽曲単位でのランダムプレイ(シャッフルプレイ)機能が可能なように、区間単位でのランダムプレイ機能も実現可能である。また試聴よりも長時間の利用が許される場面で、楽曲構造を見ながら音楽再生位置をインタラクティブに変更して聴くときには、従来の鑑賞に比べてより分析的に聴く傾向がある。例えば、楽曲全体がどのような構造になっているかを確認、構成上の各区間を聴いたり、繰り返される区間同士の比較をしたりする場面がみられる。また、楽曲構造の視覚化結果は、それを眺めながら音楽を聴くだけでも面白く有用である。

さらに楽曲中の繰り返し区間の先頭へ再生位置をジャンプするだけでなく、繰り返されない区間(間奏・ギターソロ等)にもラベリングしておけば、それらの先頭へも再生位置をジャンプできる。

次に、前述の自動サビ区間検出手法で用いるサビ区間を検出する方法の実施例を詳細に説明する。図15は、転調を伴う楽曲中のサビ区間を検出する方法の処理ステップを示すフローチャートである。

(1) 本実施例では、まず、音響信号(音響信号データ)を得る(ステップS1)。

(2) 次に、その入力音響信号の各フレームから、細部の変形の影響を受け難い12次元の特徴量(12音名各々の周波数のパワーを複数のオクターブに渡って加算した12次元クロマベクトル)を抽出する(ステップS2)。

(3) その抽出された12次元クロマベクトルの特徴量と過去の全フレームの特徴量との間の類似度を計算する(課題1に対応)(ステップS3-1)。次に

、判別基準に基づく自動閾値選定法〔先行技術 6〕によって、繰り返しの判断基準を楽曲ごとに自動的に変えながら、繰り返し区間のペアをリストアップする（課題 2 に対応）（ステップ S 3-2）。そして、それらのペアを楽曲全体に渡って統合することで、繰り返し区間のグループを作り、それぞれの端点も適切に求める（課題 3 に対応）（ステップ S 3-3）。

（4）ここで、転調を考慮に入れた場合、クロマベクトルの各次元は音名に対応しているため、その転調幅に応じて次元間で値をシフトさせた転調後のクロマベクトルと、転調前のクロマベクトルとは値が近くなる。そこで、そのように 12 種類の転調先を考慮して、転調前後のクロマベクトルの類似度を計算する。それを出発点として、上記の繰り返し区間の検出処理も 12 種類分行い、それら全ての繰り返し区間を統合する（課題 4 に対応）（ステップ S 4）。

（5）最終的に、得られた各区間のサビらしさを上記の仮定に基づいて評価する（ステップ S 5）。

（6）最もサビらしい区間の一覧を出力する（ステップ S 6）。

（7）同時に、中間結果として得られた繰り返し構造も出力する（ステップ S 7）。

また図 16 は、本発明のサビ区間を検出する装置の実施の形態の一例の構成の概略を示すブロック図である。この装置では、図 15 の方法も当然にして実現可能である。さらに図 17 は、図 16 の装置をコンピュータを利用して実現する場合に用いるプログラムのアルゴリズムの一例を示すフローチャートである。図 16 の装置の構成を説明しながら、併せて図 15 のステップと図 17 のフローチャートのステップについて説明する。

まずサンプリング手段 101 は、所定のサンプリング幅を持って重複しながらデータのサンプリングを行うハニング窓等のサンプリング技術を用いて、所定の時間単位（例えば 80 ms）で、入力されてくる音楽音響データについてサンプリングを行う（図 17 のサンプリングステップ S T 1）。データが音響信号であれば、サンプリングされるデータは、非常に短い断片（フレーム）の音響信号である。

特徴量抽出手段 103 は、サンプリング手段 101 で時間単位でサンプリング

したデータについて、音響特徴量を求める（図17の特徴量抽出ステップST2）。ここで特徴量抽出手段103で採用する音響特徴量の求め方は任意である。この実施の形態では、特徴量抽出ステップで求める音響特徴量として、1オクターブの範囲に含まれる12の音名の周波数のパワーを複数のオクターブに渡ってそれぞれ加算して得た12次元クロマベクトル（chroma vector）を用いる。

ここで図18及び図19を用いて12次元クロマベクトルについて説明する。クロマベクトルは、先行技術7に開示されているクロマ（音名，chroma）を周波数軸として、パワーの分布を表現した特徴量である。ここでクロマベクトルは、先行技術8のchroma spectrumのクロマの軸を12個の音名に離散化したものに近いものである。図18に示すように、先行技術7によれば、音楽的な音高の知覚（音楽的高さと音色的高さ）は上に昇る螺旋状の構造を持つ。そして音楽的な音高の知覚は、この螺旋を真上から見た円周上のクロマと、横から見たときの縦方向のハイト（オクターブ位置，height）の二つの次元で表現することができる。クロマベクトルでは、パワースペクトルの周波数軸がこの螺旋状の構造に沿っていると見なし、螺旋をハイト軸方向につぶして円にすることで、周波数スペクトルを円周上（1周が1オクターブ）のクロマの軸だけで表現する。つまり、異なるオクターブの同じ音名の位置のパワーを加算して、クロマ軸上のその音名の位置のパワーとする。

本実施の形態では、図19に示すように、このクロマベクトルを12次元で表し、クロマベクトルの各次元の値が平均律の異なる音名のパワーを表すものとする。図19では、6オクターブの同じ音名の位置のパワーをそれぞれ加算してクロマ軸上のその音名の位置のパワーとする状態を示している。12次元のクロマベクトルを得るためには、まず時刻tの入力音響信号に対する短時間フーリエ変換（STFT）を計算する。その後、短時間フーリエ変換（STFT）で求めた演算結果を、周波数軸を対数スケールの周波数fに変換して、パワースペクトル $\Psi_p(f, t)$ を求める。対数スケールの周波数はcentの単位で表し、Hzで表された周波数fHzを、次のようにcentで表された周波数fcentに変換する。



$$f_{cent} = 1200 \log_2 \{ f_{Hz} / (440 \times 2^{3/12-5}) \} \quad \dots (1)$$

平均律の半音は100 centに相当し、1オクターブは1200 centに相当する。そのため、音名 $c$  ( $c$ は $1 \leq c \leq 12$ の整数で、クロマに対応)、オクターブ位置 $h$  (ハイトに対応)の周波数 $F_{c,h}$ は、

$$F_{c,h} = 1200h + 100(c-1) \quad \dots (2)$$

と表せる。

この対数スケール軸のパワースペクトル $\Psi_p(f, t)$ から音名 $c$ の位置のパワーを $O_{ct_L}$ から $O_{ct_H}$  (現実の実装では、3~8)のオクターブ範囲で加算して、12次元クロマベクトルの各次元 $v_c(t)$ を下記式(3)で求める。

$$v_c(t) = \sum_{h=O_{ct_L}}^{O_{ct_H}} \int_{-\infty}^{\infty} BPF_{c,h}(f) \Psi_p(f, t) df \quad \dots (3)$$

を求める。ここで、 $BPF_{c,h}(f)$ は、音名 $c$ 、オクターブ位置 $h$ の位置のパワーを通過させるバンドパスフィルタであり、下記式(4)のように、ハニング窓の形状で定義する。

$$BPF_{c,h}(f) = \frac{1}{2} \left( 1 - \cos \frac{2\pi(f - (F_{c,h} - 100))}{200} \right) \quad \dots (4)$$

こうして得られたクロマベクトルを特徴量とすることで、繰り返す度に繰り返し区間のメロディーや伴奏が多少変わっても、繰り返し区間全体の響き (同時に鳴っている音名の構成) が類似していれば、その区間は繰り返し区間として検出できる。さらに、後述するように、類似度の工夫によって転調された繰り返し区間の検出も可能となる。

なお現在作成している装置では、音響信号を標本化周波数16 kHz、量子化ビット数16 bitでA/D変換している。そして窓関数 $h(t)$ として窓幅4

096点のハニング窓を用いた短時間フーリエ変換（STFT）を、高速フーリエ変換（FFT）で計算する。高速フーリエ変換（FFT）のフレームは、1280点ずつシフトし、すべての処理の時間単位（1フレームシフト）を80msとする。

図16に戻って、上記のようにして求めた特徴量は、特徴量記憶手段105に記憶される。そして類似度演算手段107は、それまでに入力された音楽音響データについて求めた複数の音響特徴量の相互間の類似度を求める（図17の類似度演算ステップST3）。類似度を求める際に用いる演算式は、任意であり、公知の類似度演算式のいずれを用いてもよい。そして繰り返し区間リストアップ手段109は、類似度に基づいて音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする（図17の繰り返し区間リストアップステップST4）。

類似度演算手段107では、今回求めた音響特徴量と先に求めた全ての音響特徴量との間の類似度を求めている。これによってリアルタイムにサビ区間を検出することが可能になる。具体的な類似度演算手段107では、図20及び図21に示すように、時刻 $t$ の12次元クロマベクトル（音響特徴量）とそれよりラグ $l$ （ $0 \leq l \leq t$ ）（ $l$ はアルファベット $L$ の小文字）だけ過去の全ての12次元クロマベクトルとの間の類似度を求めることになる。12次元クロマベクトル間の類似度の計算（図17のステップST3）について説明する。

時刻 $t$ の12次元クロマベクトル $\vec{v}(t)$ （但しここで $\vec{v}$ はベクトル）と、それよりラグ（lag） $l$ （ $0 \leq l \leq t$ ）だけ過去の12次元クロマベクトル $\vec{v}(t-l)$ （但しここで $\vec{v}$ はベクトル）との類似度 $r(t, l)$ を下記式（5）に基づいて求める。

$$r(t, l) = 1 - \frac{\left| \frac{\vec{v}(t)}{\max v_e(t)} - \frac{\vec{v}(t-l)}{\max v_e(t-l)} \right|}{\sqrt{12}} \quad \dots\dots\dots(5)$$

上記式(5)において、分母の $(12)^{1/2}$ は、1辺の長さがラグ1の12次元超立方体の対角線の長さであることを示している。上記式(5)中の分子中の下記式(6)は、常にその超立方体の原点を含まない面上に位置するため、 $0 \leq r(t; 1) \leq 1$ となる。

$$\frac{\vec{v}(t)}{\max_{v_i(t)} v_i(t)} \quad \dots\dots\dots (6)$$

すなわち類似度 $r(t, 1)$ は、各時刻 $t$ のクロマベクトルを最大要素で正規化し、ラグ1だけ過去のクロマベクトルとユークリッド距離を計算し、その計算結果を1から引いた値である。

次に、繰り返し区間リストアップ手段109における繰り返し区間のリストアップ(図17のステップST4)について説明する。図22は、ある楽曲に対する後述する類似線分、類似度 $r(t, 1)$ 、パラメータ空間 $R_{all}(t, 1)$ の概念図である。繰り返し区間リストアップ手段109では、図22に示すように、一方の軸を時間軸とし他方の軸をラグ軸とし、予め定めた時間長さ以上類似度が予め定めた閾値以上ある場合には、類似線分を時間軸を基準にした繰り返し区間としてリストアップする。図22においては、類似線分を時間軸と平行に表示している。なおこのリストアップは、演算上のリストアップであればよく、実際的に表示手段上にリストアップする必要はない。したがって時間軸及びラグ軸も理論上の軸であればよい。ここで「類似線分」の概念は、本願明細書において定義するものである。「類似線分」とは、予め定めた時間長さ以上類似度が予め定めた閾値以上あるときに、閾値以上ある類似度の部分の長さに対応する時間長さを有する線分として定義される。なお類似度の大きさは、類似線分に現れることはない。また閾値を適宜に変更または調整することにより、ノイズを除去することが可能になる。

図22において、類似度 $r(t, 1)$ は右下半分の三角形内で定義される。実際に得られる $r(t, 1)$ は、図23に示すように、ノイズを多く含み、サビに関連しない類似線分も存在して曖昧ことが多い。

リストアップのために、類似度 $r(t, 1)$ に基づいて、どの区間が繰り返さ

れているかを調べる。図 2 2 に示すように、類似度  $r(t, l)$  を、横軸が時間軸  $t$ 、縦軸がラグ軸  $l$  の  $t-l$  平面に描画すると、繰り返されている区間に対応して、時間軸に平行な線分（類似度が連続して高い領域）が現れる。そこで、時刻  $T_1$  から  $T_2$  の区間（以下、 $[T_1, T_2]$  と表記する）に渡ってラグ軸  $L_1$  の位置に高い類似度を持つ線分を類似線分と呼び、 $[t = [T_1, T_2], l = L_1]$  で表す。これは、 $[T_1, T_2]$  と  $[T_1 - L_1, T_2 - L_1]$  が繰り返す区間であることを意味する。よって、 $r(t, l)$  中の類似線分をすべて検出すれば、繰り返し区間の一覧が得られる。

ここで類似線分の考え方について簡単に説明する。例えば、 $t-l$  平面に図 2 4 に示すように、繰り返し区間を示す類似線分が現れている場合を考える。図 2 4 の横軸の下に示したアルファベットの表記は、それまでに入力された音響信号が A メロ → B メロ → サビ (C) → サビ (C) であることを示している。このような類似線分が現れているのは、サビ C が 2 回連続しているためである。すなわち図 2 5 に示すように、前のサビ C の区間と後のサビ C の区間との間の類似度は、最後のサビ C の区間と他の最初の二つの区間 (A, B) との類似度と比べて高くなるため、最後のサビ C に対応する時間位置で且つラグ  $l$  が前のサビ C の位置に対応する部分にサビ C と同じ時間長さの類似線分が現れるのである。さらに時間が過ぎて、図 2 6 のようになったと仮定する。図 2 6 においては、理解を容易にするために、特徴量が対比された区間を各 A, B, C のアルファベットの右下に数字で示してある。例えば「A<sub>12</sub>」の表示は、A 1 区間の A メロと A 2 区間の A メロの特徴量の類似度が演算されて、その類似度が高いために現れた類似線分であることを示している。同じく「C<sub>36</sub>」は C 3 区間のサビ区間と C 6 区間のサビ区間の特徴量の類似度が演算されて、その類似度が高いために現れた類似線分であることを示している。なお 1 つのサビ区間内においてサビの 2 度の繰り返しがある場合には、図 2 7 に示すように類似線分が現れることになる。

この時間軸に平行な線分の検出をコンピュータを用いて演算により実行するために、 $t-l$  平面における画素  $(T, L)$  ごとに  $L$  の位置に対応する 1 次元パラメータ空間  $l$  (ラグ) に画素の輝度を累積する。そして、パラメータ空間  $l$  の累積値の大きい位置のラグに、線分が存在するものとみなす。

具体的には、時刻  $t$  におけるパラメータ空間  $R_{all}(t, l)$  は、下記式 (7) から求めることができる。

$$R_{all}(t, l) = \int_1^t \frac{r(\tau, l)}{t - \tau} d\tau \quad \dots\dots\dots(7)$$

図 22 に示されるように、上記  $R_{all}(t, l)$  が大きい値を持つ  $l$  の位置に類似線分が存在する可能性が高いと考える。

なお、広帯域ノイズ等に起因する各成分がほぼ等しいクロマベクトルからは、他のクロマベクトルへの距離が比較的近くなってしまう傾向があり、 $r(t, l)$  中に類似度の高い直線（以下、ノイズ直線と呼ぶ）として現れることがある。このノイズ直線は、 $t - l$  平面において、時間軸に垂直（上下）方向、あるいは、斜め右上・左下方向に現れる。そこで、前処理として上記式 (7) の計算前にノイズ直線の抑制を行う。まず、各  $r(t, l)$  において、右、左、上、下、右上、左下の 6 方向の近傍区間の平均値を計算し、その最大値と最小値を求める。そして、右か左の方向の近傍区間の平均値が最大の場合は、類似線分の一部とみなして、強調するために  $r(t, l)$  から最小値を引く。その他の方向の近傍空間の平均値が最大の場合は、ノイズ直線の一部とみなして、抑制するために  $r(t, l)$  から最大値を引く。このようにして求めた  $R_{all}(t, l)$  は、図 28 の右側に示すような線図となる。

上記のように、 $R_{all}(t, l)$  を求めた後の類似線分の検出は、以下の手順 1 及び 2 に従って行う。

#### 手順 1：線分候補ピークの検出

図 28 の右側の線図に示される  $R_{all}(t, l)$  中の十分に高いピークを、線分候補ピークとして検出する。まず、 $R_{all}(t, l)$  の  $lag$  軸方向のピークを、2 次多項式適合による平滑化微分を用いたピーク検出〔先行技術 9〕により求める。具体的には、下記式 (8) で求める  $R_{all}(t, l)$  の平滑化微分が正から負に変わる箇所をピークとする ( $KSize = 0.32_{sec}$ )。

$$\sum_{w=-KSize}^{KSize} w R_{all}(t, l+w) \dots\dots\dots (8)$$

ただし、このピーク検出の前に、 $R_{all}(t, l)$ のlag軸方向に、2階のカーディナルBースプライン関数を重み関数とする移動平均によってスムージングをかけたものを引いて、 $r(t, l)$ のノイズ成分等の蓄積による大局的な変動を取り除いておく〔 $R_{all}(t, l)$ にハイパスフィルタをかけることに相当する〕。

次に、こうして得られたピークの集合から、ある閾値より大きいピークのみを、線分候補ピークとして選ぶ。前述の課題2で述べたように、この閾値は楽曲ごとに適切な値が異なるため、楽曲に基づいて自動的に変える必要がある。そこで、 $R_{all}(t, l)$ のピーク値を閾値によって二つのクラスに分けるときに、クラス分離度を最大とする判別基準に基づく自動閾値選定法〔先行技術6〕を用いる。この自動閾値選定法は、図29に示すように閾値によって二つのクラスに分けるという考え方を採用している。ここでは、クラス分離度としてクラス間分散

$$\sigma^2_B = \omega_1 \omega_2 (\mu_1 - \mu_2)^2 \dots\dots\dots (9)$$

を最大とする閾値を求める。ただし、 $\omega_1 \omega_2$ は、閾値によって分けられた二つのクラスの生起確率（各クラスのピーク個数／全体のピーク個数）、 $\mu_1$ 、 $\mu_2$ は、各クラスのピーク値の平均である。

#### 手順2：類似線分の探索

図30に示すように、各線分候補ピークのlag軸上の位置lにおいて、類似度 $r(t, l)$ の時間軸方向を一次元関数とみなして、それが連続して十分高い区間を探索し、類似線分とする。

まず、 $r(t, l)$ の時間軸方向に、2階のカーディナルBースプライン関数を重み関数とする移動平均によってスムージングをかけた $r_{smooth}(t, l)$ を求める。次に、 $r_{smooth}(t, l)$ 中で、ある閾値を連続して越えているすべての区間のうち、一定の長さ（6.4sec）以上のものを類似線分として求める

。この閾値も、上記の判別基準に基づく自動閾値選定法により定める。ただし、今度はピーク値を扱うのではなく、ピーク値が高い上位5個の線分候補ピークを選び、それらのラグ1の位置の  $r_{smooth}(\tau, 1)$  ( $1 \leq \tau \leq t$ ) がとる値を二つのクラスに分ける。

上記のようにしてリストアップされた繰り返し区間のリストは、図16に示すリスト記憶手段111に記憶される。統合繰り返し区間決定手段113は、リスト記憶手段111に記憶されたリストから複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある1以上の繰り返し区間を統合して一つの統合繰り返し区間を決定する。そして統合繰り返し区間決定手段113は、さらに決定した複数の統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する。

この統合繰り返し区間決定ステップ（図17のST5）では、図31に示すように、前述の  $t-1$  平面における時間軸の共通区間に存在するリストアップした類似線分どうしをそれぞれグルーピングにより統合して統合繰り返し区間RPと定める。そして複数の統合繰り返し区間RPを、共通区間の位置及び長さでグルーピングされる類似線分のラグ軸で見た位置関係とに基づいて複数種類の統合繰り返し区間列に分類する。

より具体的には、図31に示すように、リストアップされた複数の繰り返し区間  $C_{12} \sim C_{56}$ （類似線分）の相互関係は、時間軸上の共通区間に対応する過去のラグ位置に1以上の繰り返し区間  $C_{12} \sim C_{56}$ （類似線分）が存在するか否かと、そのラグ位置に対応する過去の時間帯において繰り返し区間（類似線分）が存在するか否かの関係である。例えば、C6の共通区間に繰り返し区間を示す類似線分  $C_{16}$  がある場合、その繰り返し区間のラグ位置に対応する過去のラグ位置にも類似線分  $C_{12}$  があるという関係である。これらの関係に基づいて、このステップでは、共通区間に対応する過去のラグ位置に1以上の繰り返し区間（類似線分）がある場合に、それらをグルーピング化してその共通区間に繰り返し区間（類似線分）があるものと決定し、その繰り返し区間を統合繰り返し区間RP2, RP5, RP6等とする。ただし、図32に示すように、本来存在している最初の繰り返し区間に対応しては、過去の時間帯には類似線分は無い。そのため最初の繰り返し区間に対応する統合繰り返し区間RP1については、最初の統合繰り返し

区間RP2とその共通区間に存在する類似線分 $C_{12}$ を基準にして補足する。なおこの補足は、プログラミングによって簡単に実現できる。このようにして1種類の統合繰り返し区間列が作られる。

図33は、共通区間の長さが長い場合の統合繰り返し区間RP1及びRP2の列を作る場合の状況を示している。図34は、図27のようにサビ区間に2回の繰り返しがあるために、統合繰り返し区間RPの共通区間の長さが図31及び図32の統合繰り返し区間列を構成する統合繰り返し区間の $1/2$ になる場合の状況を示している。このようにして統合繰り返し区間決定ステップでは、決定した複数の統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する。この分類化は、共通区間の長さの共通性と、共通区間に存在する繰り返し区間（類似線分）の位置関係と数との関係に基づいて行われる。

統合繰り返し区間決定手段113により決定した、統合繰り返し区間は統合繰り返し区間列として統合繰り返し区間記憶手段115に記憶される。図35は、統合繰り返し区間列を表示手段118に表示した一例を示している。

前述の統合繰り返し区間決定手段113で実行されている統合処理をコンピュータを用いてより高い精度で実行する場合のより具体的な手順について説明する。前述の各類似線分は、ある区間が二回繰り返されていることだけを表すため、例えばAとA'のペア、A'とA''のペアが、それぞれ繰り返し区間として検出されたときには、それらを一つの繰り返し区間のグループとして統合する必要がある。ここで、ある区間がn回( $n \geq 3$ )繰り返されている場合には、もれなく検出されるとすると、 $n(n-1)/2$ 本の類似線分が検出される。そこで、同じ区間の繰り返しを表す類似線分をグルーピングし、繰り返し区間を統合する。さらに、もれていた類似線分の検出や、得られた類似線分が適切であるかの検証も行う。

この統合処理は、以下の手順で実現する。

手順1：類似線分のグルーピング

ほぼ同じ区間の類似線分を、一つのグループにまとめる。各グループ $\phi_i = [ [T_{si}, T_{ei}], T_i ]$ は、区間 $[T_{si}, T_{ei}]$ と、類似線分（区間が決まれば、線分候補ピークと対応する）のlag値 $v_{ij}$ の集合 $T_i = \{v_{ij} \mid j = 1, 2$



, ...,  $M_i$  } ( $M_i$  はピークの個数) で表される。そして、この類似線分のグループ  $\phi_i$  の集合を、 $\Phi = \{ \phi_i \mid i = 1, 2, \dots, N \}$  ( $N$  はグループの個数) とする。

#### 手順2：線分候補ピークの再検出

グループ  $\phi_i$  ごとに、区間  $[T_{s_i}, T_{e_i}]$  内の類似度  $r(t, l)$  に基づいて、類似線分を改めて求めなおす。これにより、もれていた類似線分の検出ができ、例えば、図22で、ABCCの繰り返しに相当する長い類似線分上で、Cの繰り返しに相当する類似線分2か所が得られていなくても、この処理で検出されることが期待できる。

まず、 $[T_{s_i}, T_{e_i}]$  内に限定して、パラメータ空間  $R_{[T_{s_i}, T_{e_i}]}(l)$  ( $0 \leq l \leq T_{s_i}$ ) を下記式 (10) で作成する。

$$R_{[T_{s_i}, T_{e_i}]}(l) = \int_{T_{s_i}}^{T_{e_i}} \frac{r(r, l)}{T_{e_i} - T_{s_i}} dr \quad \dots\dots\dots(10)$$

次に、前述の線分候補ピークの検出と同様に、平滑化微分を用いたピーク検出を行い ( $K_{Size} = 2, 8_{sec}$ )、自動閾値選定法で定めた閾値を越えた線分候補ピークの  $lag$  値  $v_{ij}$  の集合を、改めて  $T_i$  とする。

自動閾値選定法では、 $\Phi$  の全グループの区間における  $R_{[T_{s_i}, T_{e_i}]}(l)$  のピーク値を、二つのクラスに分けるようにする。

#### 手順3：類似線分の適切さの検証1

サビと無関係な類似線分からなるグループ  $\phi_i$ 、あるいは、 $T_i$  の中で無関係な線分と考えられるピークを削除する。

似た伴奏の繰り返しが多用される楽曲の場合サビと関係ない線分候補ピークが  $R_{[T_{s_i}, T_{e_i}]}(l)$  に等間隔に多く現れる傾向がある。

そこで、 $R_{[T_{s_i}, T_{e_i}]}(l)$  に対して平滑化微分を用いたピーク検出を行い、一定間隔 (間隔は任意) で連続して並ぶ高いピークの個数が10個より多いときサビと無関係な類似線分からなるグループだと判断し、そのグループを  $\Phi$  から削除

する。

また、一定間隔で連続して並ぶ低いピークの個数が5個より多いとき、サビと無関係な線分候補ピークだと判断し、その一連のピークを $T_i$ から削除する。

#### 手順4：類似線分の適切さの検証2

$T_i$ の中には、区間  $[T_{s_i}, T_{e_i}]$  の一部分だけ類似度が高いピークが含まれることがあるため、そうした類似度の変動の大きいピークを削除する。そこで、当該区間の  $r_{smooth}(\tau, l)$  の標準偏差を求め、ある閾値より大きいものは $T_i$ から削除する。この閾値は、 $\phi_i$ の中で、上記で求めた類似線分に対応する線分候補ピークは信頼できると考え、それらのピークでの上記標準偏差の最大値を定数倍（1.4倍）して定める。

#### 手順5：類似線分の間隔の考慮

繰り返し区間が重ならないようにするために、lag軸上で隣接する類似線分（線分候補ピーク）の間隔を、線分の長さ $T_{e_i} - T_{s_i}$ 以上とする必要がある。そこで、線分の長さより狭い間隔を持つ二つのピークのいずれかを、全体として高いピーク集合が残るように削除し、すべての間隔が類似線分の長さ以上になるようにする。

#### 手順6：共通区間を持つグループを統合

$T_i$ の各ピークについて、そのlag値 $u_{ij}$ だけの過去の区間  $[T_{s_i} - u_{ij}, T_{e_i} - u_{ij}]$  のグループがあるかを探索し、発見したら統合する。統合処理では、発見したグループのすべてのピークを、対応するlag値の場所に持つように、 $T_i$ に線分候補ピークを追加する。発見したグループ自体は削除する。

さらに、区間  $[T_{s_i} - u_{ij}, T_{e_i} - u_{ij}]$  に一致する線分候補ピークを持つグループ $T_k$ （グループの区間自体は異なる）があるかも探索し、発見したら統合するか判断する。この場合、 $T_k$ の過半数のピークが $T_i$ に含まれていれば、上記同様の統合処理を行う。含まれていなければ、 $T_i$ と $T_k$ で同じ区間を指しているピークを比較し、低い方を削除する。上記で実際に統合がなされたら、後処理として手順5の処理を再び行う。

次に、転調を伴う繰り返しの検出（図15のステップS4）について説明する。以上述べてきた処理は転調を考慮していなかった。しかし上記の処理は、以下

のように転調を扱える処理へと容易に拡張できる。図 3 6 に示すように、転調前と転調後の 1 2 次元クロマベクトルは異なる。そこで特徴量抽出ステップ（図 1 5 のステップ S 2）では、図 3 7 に示すように、1 2 次元クロマベクトルからなる音響特徴量を 1 転調幅ずつ 1 1 転調幅までシフトして得た転調幅の異なる 1 2 種類の音響特徴量を求める。次に類似度演算ステップ（図 1 5 のステップ S 3 - 1）では、今回求めた音響特徴量と先に求めた全ての 1 2 種類の音響特徴量との間の類似度を、時刻  $t$  の今回の音響特徴量を表す 1 2 次元クロマベクトルとそれよりラグ  $l$  ( $0 \leq l \leq t$ ) だけ過去の全ての 1 2 種類の音響特徴量を表す 1 2 次元クロマベクトルとの間の類似度として演算する。そして繰り返し区間リストアップステップ（図 1 5 のステップ S 3 - 2）では、図 3 8 に示すように、1 2 種類の音響特徴量ごとに、一方の軸を時間軸  $t$  とし他方の軸をラグ  $l$  とし、予め定めた時間長さ以上類似度が予め定めた閾値以上ある場合には類似度が予め定めた閾値以上である部分の長さに対応する時間長さを有する類似線分を時間軸を基準にした繰り返し区間としてそれぞれ 1 2 種類のリストをリストアップする。統合繰り返し区間決定ステップ（図 1 5 のステップ S 3 - 3 及び S 4）では、1 2 種類のリストごとに、時間軸の共通区間に存在するリストアップした類似線分どうしをそれぞれグルーピングにより統合して統合繰り返し区間と定める（S 3 - 3）。さらに 1 2 種類のリストについて定めた複数の統合繰り返し区間を共通区間の時間軸上の存在位置及び長さ、グルーピングされる類似線分のラグ軸で見た位置関係とに基づいて複数種類の転調を考慮した複数種類の統合繰り返し区間列に分類化する（S 4）。このようにすると、転調を含んだ音楽音響データであっても、転調した部分の特徴量を 1 1 段階の転調幅のシフトでずらして類似度を求めるため、転調した部分の特徴量を正しく抽出することができる。

楽曲が転調を含んでいる場合に、これをコンピュータを用いてより具体的に処理する場合には、上記の処理を以下のとおりにする。ここで、転調は平均律の半音  $t_r$  個分上の調へ変わることによって表すことにする。 $t_r$  は 0, 1, ..., 11 の 1 2 種類の値を取るものとする。 $t_r = 0$  は転調しないことを意味し、 $t_r = 10$  は半音 10 個分上か、全音分下へ転調することを意味する。

1 2 次元クロマベクトル  $v(t)$ （ここで  $v$  はベクトル）は、各次元  $v_c(t)$

の値を次元間で  $t_r$  個分だけシフトさせることで、転調を表現できる特長を持つ。具体的には、ある演奏の 12 次元クロマベクトルを  $v(t)$ （ここで  $v$  はベクトル）とし、それを  $t_r$  個上へ転調した演奏の 12 次元クロマベクトルを  $v(t)$ （ここで  $v$  はベクトル）とすると、

$$v(t) \approx S^{t_r} v(t) \quad \dots (11)$$

となる。

ただし、 $S$  はシフト行列で、下記式 (12) のように 12 次正方行列を一つ右にシフトした行列として定義される。

$$S = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & 1 & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 \\ 1 & 0 & \dots & \dots & \dots & 0 \end{pmatrix} \quad \dots (12)$$

転調を伴う繰り返しの検出の処理手順を以下に述べる。まず、クロマベクトルのこの特長を利用し、 $t_r$  ごとの 12 種類の類似度  $r_{tr}(t, 1)$  を下記式 (13) と定義しなおす。

$$r_{tr}(t, 1) = 1 - \frac{\left| \frac{S^{t_r} \vec{v}(t)}{\max_c v_c(t)} - \frac{\vec{v}(t-1)}{\max_c v_c(t-1)} \right|}{\sqrt{12}} \quad \dots (13)$$

次に、それぞれの類似度  $r_{tr}(t, 1)$  に対して、前述した繰り返し区間のリストアップをする。ただし、自動閾値選定法は  $t_r = 0$  のときだけ適用し、他の  $t_r$  では、 $t_r = 0$  で定めた閾値を用いる。これにより、転調のない曲で、 $t_r = 0$  以外のときに類似線分が誤検出されにくくなる。そして、こうして得られた各  $t_r$  ごとの類似度と類似線分に対して、前述の統合処理を行う。その結果、 $t_r$  ごとに別々の類似線分のグループ  $\phi_{tr, i}$  の集合  $\Phi_{tr}$  が得られる。そこで前述した、共通区間を持つグループの統合の処理を、 $t_r$  間にまたがって行う（異な

る  $t_r$  に対して共通区間を持つグループを探索する) ことで、転調を含む繰り返し区間を一つのグループとして統合する。ただし、前出の処理では「 $T_k$  の過半数のピークが  $T_l$  に含まれていれば、上記同様の統合処理を行う」とあるが、ここでは常に統合処理を行う。

以下、異なる  $t_r$  から得られたグループも合わせて、 $\Phi = \{\phi_i\}$  で表す。転調区間が後から分かるように、どの  $t_r$  から統合されたかという情報は保存しておく。

図 16 に戻って、サビ区間決定手段 117 では、統合繰り返し区間記憶手段 115 に記憶された統合繰り返し区間列からサビ区間を決定する。なお図 16 の例では、サビ区間を含む統合繰り返し区間列または複数種類の統合繰り返し区間列は、表示手段 118 に表示される (図 41 参照)。そしてサビ区間を含む統合繰り返し区間列が他の統合繰り返し区間列とは異なる表示態様で表示される。このようにすると検出したサビ区間を他の繰り返し区間とは区別して明瞭に表示することができる。なおこの例では、統合繰り返し区間列を、表示手段 118 に表示させながら選択手段 121 で選択して、音響の再生手段 123 でサビ区間を含む統合繰り返し区間列またはその他の統合繰り返し区間列を選択的に再生することができる。

図 15 及び図 17 のサビ区間決定ステップ (S5、ST6) では、例えば、統合繰り返し区間列に含まれる統合繰り返し区間の類似度の平均と、統合繰り返し区間の数と長さとに基づいて統合繰り返し区間列に含まれる統合繰り返し区間のサビらしさを求める。そして、最もサビらしさの高い統合繰り返し区間列に含まれる統合繰り返し区間をサビ区間として選択する。最初に図 39 及び図 40 を用いて説明した前述の仮定 1 乃至仮定 3 を満たす統合繰り返し区間は、一般的にはサビらしさが高い。

上記の仮定を考慮して、コンピュータを用いてサビ区間を自動的に選択する方法について以下に説明する。前述の類似線分のグループの集合  $\Phi$  の中から、ある一つのグループをサビ区間として選ぶ。そのために、各グループ  $\phi_i$  のサビらしさ  $v_i$  を、類似線分の平均類似度や上記した仮定に基づいて評価し、最もサビらしさ  $v_i$  の高いグループをサビ区間であると判定する。その準備として、グループごと

に、類似線分（線分候補ピーク $v_{ij}$ ）をそれが指す二つの区間へ展開し、すべての繰り返し区間 $[P s_{ij}, P e_{ij}]$ とその信頼度 $\lambda_{ij}$ のペアの集合を下記式（14）により求める。

$$\Lambda_i = \{ [P s_{ij}, P e_{ij}], \lambda_{ij} \mid j = 1, 2, \dots, M_i + 1 \} \quad \dots (14)$$

ここで、 $[P s_{ij}, P e_{ij}] = [T s_{i-v_{ij}}, T e_{i-v_{ij}}]$  とし、信頼度 $\lambda_{ij}$ は、対応する類似線分における類似度 $r_{i-v_{ij}}(t, 1)$ の平均とする。ただし、 $j = M_i + 1$ のときは、下記式（15）のようになる。

$$[P s_{ij}, P e_{ij}] = [T s_i, T e_i], \quad \lambda_{ij} = \max_{k=1}^{M_i} \lambda_{ik} \quad \dots (15)$$

サビらしさ $v_{ij}$ は、以下の手順で評価する。

（1）仮定2を満たす統合繰り返し区間の信頼度を増加

仮定2で述べたAメロ～サビに相当するような十分に長い統合繰り返し区間（50 sec以上）を持つグループ（統合繰り返し区間列） $\phi_n$ に関して、その各区間の終了点 $P e_{nk}$ とほぼ等しい終了点 $P e_{ij}$ を持つ区間が他のグループ（他の統合繰り返し区間列）にあるか探索する。発見されれば、発見されたその統合繰り返し区間がサビである可能性が高いと考え、その信頼度 $\lambda_{ij}$ を2倍する。

（2）仮定3を満たす統合繰り返し区間の信頼度を増加

サビとして適切な区間長の範囲（仮定1）の統合繰り返し区間 $[P s_{ij}, P e_{ij}]$ に関して、その区間の半分程度の短い統合繰り返し区間が前半と後半に一つずつ存在するか調べる。存在する場合には、それら二つの区間の信頼度の平均の半分を、元の区間の信頼度 $\lambda_{ij}$ に加える。

（3）サビらしさを算出

上記で得られた信頼度に基づき、サビらしさを下記式（16）で算出する。

$$v_{ij} = \left( \sum_{j=1}^{M_i+1} \lambda_{ij} \right) \cdot \log \frac{T e_i - T s_i}{D_{i,0}} \quad \dots (16)$$

上記式(16)において、 $\Sigma$ の項は、グループ(統合繰り返し区間列) $\phi_i$ 中にある統合繰り返し区間の数が多いほど、また、それらの信頼度が高いほど、サビらしさが高いことを意味する。 $\log$ の項は、そのグループ(統合繰り返し区間列)に含まれる統合繰り返し区間が長いほど、サビらしさが高いことを意味する。定数 $D_{len}$ は予備実験の結果から1.4 secとした。

最終的に、サビとして適切な区間長の範囲(仮定1)を持つグループの中で、下記式(17)によって決まる集合 $\Lambda_m$ 中の区間 $[P_{smj}, P_{emj}]$ を、サビ区間とする。

$$m = \underset{i}{\operatorname{argmax}} v_i \quad \dots\dots\dots (17)$$

ここで後処理として、隣接する $P_{smj}$ の最小間隔を求め、区間長が最小間隔となるように $P_{emj}$ を移動して各区間を広げ、隙間を埋める。これは、本来はサビ区間が連続して隙間がないにも関わらず、得られた繰り返し区間では隙間が空いてしまうことがあるからである。ただし、埋める隙間が大きすぎるとき(1.2 sec以上で区間長の半分より広いとき)は埋めない。

図17に示すように、上記のようにサビ区間を決定したら(ステップST6)、その結果を図16の表示手段118にリアルタイムで表示する(ステップST7)。そして、音楽音響データの全データについて上記の処理が終了するまで、上記処理が繰り返される(ステップST8)。

次に、上記実施の形態のサビ区間検出装置の実際とこの装置を用いた実験結果について説明する。実験では、音楽音響信号を音楽音響データとして入力した。そして検出したサビ区間の一覧をリアルタイムに出力することとした。装置は、刻一刻と、過去の音響信号中でサビ区間と考えられる区間の一覧(リスト)を求め、中間結果として得られた繰り返し構造(繰り返し区間の一覧 $\Lambda_i$ )と共に出力し続ける。この出力を視覚化した例を図41に示す。図41において、横軸は時間軸(sec)で楽曲全体を表示しており、上半分がパワー変化、下半分の最上段がサビ区間を含む統合繰り返し区間列の一览(最後のサビは転調を伴う)、下5段が他の統合繰り返し区間列の繰り返し構造を表す。

評価実験として、「RWC研究用音楽データベース：ポピュラー音楽」〔先行

技術10]の100曲(RWC-MDB-P-2001, No. 1~100)を対象に、本装置のサビ検出性能を調べた。1曲すべてを入力し終わった時点で、サビ区間として検出されたものを対象に評価する。この正誤を判定するためには、基準となる正解のサビ区間を人間が手作業で指定する必要がある。そこで、楽曲を分割して各部にサビ、Aメロ、Bメロ、間奏等をラベリングできる、楽曲構造ラベリング用エディタを開発した。ラベリングでは、相対的な調の移動幅(曲の先頭の調に対して半音何個分上か)も正解に付与する。

こうして作成した正解に基づき、各曲に対する出力結果の区間と正解のサビ区間がどれくらい重なっているかを、再現率(recall rate)、適合率(precision rate)、および両者を統合したF値(F-measure)〔先行技術11〕の観点から評価した。以下に定義を示す。

再現率(R) = 正しく検出したサビ区間の長さの合計 / 正解のサビ区間の長さの合計

適合率(P) = 正しく検出したサビ区間の長さの合計 / 検出した区間の長さの合計

$$F \text{ 値} = (\beta^2 + 1) PR / (\beta^2 P + R) \quad (\beta = 1 \text{ を使用})$$

ただし、転調を伴う場合には、相対的な調の移動幅が正解と一致したときだけ、正しく検出したと判断した。そして、F値が0.75以上のとき、その曲のサビ区間を正しく得られた(正答した)と判定した。

評価結果として、100曲中の正答曲数を表1に示す。

表1

	条件 (使用: ○, 未使用: ×)			
転調区間の検出 仮定2,3の使用	○ ○	× ○	○ ×	× ×
正答曲数	80曲	74曲	72曲	68曲

本装置の性能は一番左の80曲(80曲の平均F値は0.938)である。誤検出は、サビの繰り返しが他の箇所の繰り返しより多くなかったり、曲中ほとん



どが類似伴奏の繰り返しだったりしたのが主な原因だった。100曲中には、サビに転調のある曲が10曲含まれているが、そのうち9曲は検出できていた。前述の転調を伴う繰り返しの検出をやめた場合、左から二番目のように性能が落ちた。一方、仮定2、3に基づく信頼度の増加をやめた場合は、右二つのようにさらに性能が落ちた。サビの繰り返しで伴奏やメロディーに大幅な変化を伴う曲は22曲あったが、そのうち21曲は検出できており、その中で変化を伴うサビ自体は16曲で検出できていた。

本発明は、基本的に楽曲中で最も多く繰り返される区間をサビとして検出する。その際、様々な区間の繰り返しを楽曲全体の情報を統合しながら調べることで、従来実現されていなかった、すべてのサビ区間の開始点・終了点の一覧を得ることを可能にした。また、転調後でも繰り返しと判断できるような、クロマベクトル間の類似度を導入したことで、サビの転調も検出できるようになった。RWC研究用音楽データベース(RWC-MDB-P-2001)100曲を用いて評価した結果、80曲正答でき、実世界の音響信号中のサビ区間が検出できることが確認された。

なお、本発明は音楽要約〔先行技術12〕とも関連しており、本発明の装置を楽曲の要約結果としてサビ区間を提示する音楽要約方法と捉えることもできる。さらに、サビ区間よりも長い区間の要約が必要なときには、中間結果として得られた繰り返し構造を用いることで、楽曲全体の冗長性を減らした要約の提示も可能となる。例えば、中間結果として(Aメロ→Bメロ→サビ)の繰り返しが捉えられているときは、それを提示できる。

この実験では、ポピュラー音楽を用いて評価したが、本発明は他の音楽ジャンルにも適用できる可能性を持つ。実際に、数曲のクラシック音楽に適用したところ、その楽曲で最も代表的な主題が提示される部分を求めることができた。

なお、本発明は上記実施例に限定されるものではなく、本発明の趣旨に基づいて種々の変形が可能であり、これらを本発明の範囲から排除するものではない。例えば、音響特徴量として、クロマベクトル以外に、周波数スペクトル、MFCC(Mel-Frequency Cepstrum Coefficients)等を用いてもよい。それらの微分値もさらに音響特徴量として加えることも

可能である。また、音響特徴量間の類似度として、以下の三つ等も考えられる。

$$r(t, l) = 1 - \frac{\left| \frac{\bar{v}(t)}{\sum_c v_c(t)} - \frac{\bar{v}(t-l)}{\sum_c v_c(t-l)} \right|}{\sqrt{2}}$$

$$r(t, l) = 1 - \frac{\left| \frac{\bar{v}(t)}{|\bar{v}(t)|} - \frac{\bar{v}(t-l)}{|\bar{v}(t-l)|} \right|}{\sqrt{2}}$$

$$r(t, l) = \frac{\bar{v}(t) \cdot \bar{v}(t-l)}{|\bar{v}(t)| |\bar{v}(t-l)|}$$

さらに、本発明は入力を音響信号でなくMIDI信号とする場合にも適用でき、その場合には、音響特徴量の代わりにMIDI信号もしくはMIDI信号特徴量を用い、類似度としてはそれらのMIDI信号もしくはMIDI信号特徴量間の距離に基づく類似度を用いればよい。MIDI信号特徴量として、12次元クロマベクトルを用いてもよいのは勿論である。その場合、MIDI信号から12次元クロマベクトルを求める任意の方法が利用できる。例えば、MIDIノートオン・オフのメッセージのオクターブ情報を消して12次元クロマベクトルを求めることができる。

以上、詳細に説明したように、本発明の実施例によれば、音楽CD(compact disc)等による実世界の複雑な混合音からサビ区間を検出し、各サビの区間の開始点と終了点の一覧を求めることができるだけでなく、転調を伴うサビ区間を検出することも可能である。その際、楽曲全体の中に含まれる様々な繰り返し構造(複数の統合繰り返し区間列)に基づいてサビ区間を検出する。さらに、楽曲全体の中に含まれる様々な繰り返し構造に基づいてサビを検出するた

め、その中間結果として、繰り返し構造の一覧も同時に得ることができる。

#### 産業上の利用可能性

本発明によれば、インターフェースによる選択により簡単にサビ区間等の楽曲構造特徴区間を音楽音響データ再生装置を用いて再生することができる利点を得られる。

本発明を用いれば、「サビ出し」機能（サビ区間選択ボタン）を搭載した音楽試聴機を得ることができる。試聴者はこのサビ区間選択ボタンを押すだけで、サビ区間の先頭へ再生をジャンプさせる（瞬時に早送りする）ことができ、自分でサビ区間を探す煩わしい作業から解放される利点がある。さらに、サビ以外の楽曲中の繰り返し区間も事前に推定あるいは用意しておくことで、次の楽曲構造の繰り返し区間の先頭へも再生をジャンプさせることができる利点を得られる。

また本発明によれば、楽曲中に出現するサビ区間を網羅的に検出することができる。また本発明によれば、1つのサビ区間がどこからどこまでかを検出することができる。さらに本発明によれば、転調されたサビ区間も検出できる。また本発明によれば、サビ区間だけでなく、その他の繰り返し区間も再生し且つ表示手段にそれぞれ表示することが可能である。

## 請求の範囲

1. 音楽音響データを再生する音楽音響データ再生装置を用いて前記音楽音響データを再生する楽曲再生方法であって、

再生の対象となる前記音楽音響データに関して、楽曲中の楽曲構造特徴区間を指定するための複数の楽曲構造特徴区間指定データを予め用意する第1のステップと、

前記複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するインターフェースを用意する第2のステップと、

前記音楽音響データ再生装置を用いて前記音楽音響データの再生を開始した後、操作者が前記インターフェースを用いて選択した前記任意の楽曲構造特徴区間に対応する前記楽曲構造特徴区間指定データに基づいて、前記音楽音響データ再生装置による前記音楽音響データの再生位置を前記任意の楽曲構造特徴区間に変更する第3のステップとからなることを特徴とする楽曲再生方法。

2. 音楽音響データを再生する音楽音響データ再生装置を用いて、楽曲中の複数の楽曲構造特徴区間から選択した任意の楽曲構造特徴区間を再生する方法であって、

再生の対象となる前記音楽音響データに関して、前記複数の楽曲構造特徴区間を指定するための複数の楽曲構造特徴区間指定データを予め用意する第1のステップと、

前記複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するインターフェースを用意する第2のステップと、

操作者が前記インターフェースを用いて選択した前記任意の楽曲構造特徴区間に対応する前記楽曲構造特徴区間指定データに基づいて前記音楽音響データ再生装置に前記任意の楽曲構造特徴区間を再生させる第3のステップとからなることを特徴とする楽曲再生方法。

3. 前記楽曲構造特徴区間が、前記楽曲中のサビ区間であることを特徴とする請求項1または2に記載の楽曲再生方法。

4. 前記楽曲構造特徴区間が、前記楽曲中の繰り返し区間を含んでいること

を特徴とする請求項 1 または 2 に記載の楽曲再生方法。

5. 前記第 1 のステップでは、前記音楽音響データに基づいて前記複数の楽曲構造特徴区間を人が特定し、特定した前記複数の楽曲構造特徴区間を指定するための前記複数の楽曲構造特徴区間指定データを作成することを特徴とする請求項 1 または 2 に記載の楽曲再生方法。

6. 前記第 1 のステップでは、対象とする前記音楽音響データを複数の試聴者が再生した際の再生行動の統計データに基いて前記複数の楽曲構造特徴区間を特定し、特定した前記複数の楽曲構造特徴区間を指定するための前記複数の楽曲構造特徴区間指定データを作成することを特徴とする請求項 1 または 2 に記載の楽曲再生方法。

7. 前記第 2 のステップで用意する前記インターフェースは、前記音楽音響データの長さを時間軸を用いた画像表示により表示し、併せて前記音楽音響データ再生装置が前記音楽音響データを再生している前記音楽音響データの再生位置と前記複数の楽曲構造特徴区間とを前記時間軸を用いた画像表示により表示する画像表示部を有していることを特徴とする請求項 1 または 2 に記載の楽曲再生方法。

8. 前記複数の楽曲構造特徴区間は、楽曲のサビ部分を示す複数のサビ区間と、繰り返し部分を示す複数の繰り返し区間とからなり、

前記画像表示部は前記複数のサビ区間と前記複数の繰り返し区間とを区別して表示することを特徴とする請求項 7 に記載の楽曲再生方法。

9. 前記インターフェースは、前記複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するために操作者が操作する 1 以上の選択ボタンを有している請求項 1 または 2 に記載の楽曲再生方法。

10. 前記 1 以上の選択ボタンには、前記複数の楽曲構造特徴区間から前記サビ区間だけを選択する 1 以上のサビ区間選択ボタンが含まれていることを特徴とする請求項 9 に記載の楽曲再生方法。

11. 前記 1 以上のサビ区間選択ボタンは前のサビ区間を選択する第 1 のサビ区間選択ボタンと、後のサビ区間を選択する第 2 のサビ区間選択ボタンとからなる請求項 10 に記載の楽曲再生方法。

12. 前記1以上の選択ボタンには、前記複数の繰り返し区間を選択する際に前の繰り返し区間を選択する第1の繰り返し区間選択ボタンと、後の繰り返し区間を選択する第2の繰り返し区間選択ボタンが含まれていることを特徴とする請求項9に記載の楽曲再生方法。

13. 前記第2のステップで用意する前記インターフェースは、前記音楽音響データの長さを時間軸を用いた画像表示により表示し、併せて前記音楽音響データ再生装置が前記音楽音響データを再生している前記音楽音響データの再生位置と前記複数の楽曲構造特徴区間とを前記時間軸を用いた画像表示により表示する画像表示部を有しており、

前記インターフェースは、前記複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するために操作者が操作する1以上の選択ボタンを有しており、

前記1以上の選択ボタンは前記画像表示部に表示された画像表示ボタンである請求項1または2に記載の楽曲再生方法。

14. 前記複数の楽曲構造特徴区間は、楽曲のサビ部分を示す複数のサビ区間と、繰り返し区間を示す複数の繰り返し区間とからなり、

前記画像表示部は前記複数のサビ区間と前記複数の繰り返し区間とを区別して表示し、

前記画像表示部に示された前記複数のサビ区間及び前記複数の繰り返し区間の表示を、前記複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するために操作者が操作する1以上の選択ボタンとすることを特徴とする請求項13に記載の楽曲再生方法。

15. 音楽音響データを再生する音楽音響データ再生装置を有し、且つ楽曲中の複数の楽曲構造特徴区間から選択した任意の楽曲構造特徴区間を再生する機能を備えた楽曲再生装置であって、

前記複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択する機能を有するインターフェースと、

再生する前記音楽音響データに対応して予め定められた前記複数の楽曲構造特徴区間を指定するための楽曲構造特徴区間指定データを格納する楽曲構造特徴区間指定データ格納手段と、

前記インターフェースによって選択された前記任意の楽曲構造特徴区間を指定する前記楽曲構造特徴区間指定データを前記音楽音響データ再生装置に与える指定データ付与手段とを具備し、

前記音楽音響データ再生装置は前記指定データ付与手段から付与された前記楽曲構造特徴区間指定データによって指定された前記楽曲構造特徴区間に再生位置を変更するように構成されていることを特徴とする楽曲再生装置。

16. 前記音楽音響データには前記楽曲構造特徴区間指定データが付随しており、

前記楽曲構造特徴区間指定データ格納手段は前記音楽音響データに付随する前記楽曲構造特徴区間指定データを読み取って格納することを特徴とする請求項15に記載の楽曲再生装置。

17. 前記音楽音響データから自動的に前記楽曲中の前記複数の楽曲構造特徴区間を抽出して前記楽曲構造特徴区間指定データを作成する特徴区間指定データ生成手段を更に備えていることを特徴とする請求項15に記載の楽曲再生装置。

18. 前記楽曲構造特徴区間が、前記楽曲中のサビ区間であることを特徴とする請求項15に記載の楽曲再生装置。

19. 前記楽曲構造特徴区間が、前記楽曲中の繰り返し区間を含んでいることを特徴とする請求項18に記載の楽曲再生装置。

20. 前記インターフェースは、前記音楽音響データの長さを時間軸を用いた画像表示により表示し、併せて前記音楽音響データ再生装置が前記音楽音響データを再生している前記音楽音響データの再生位置と前記複数の楽曲構造特徴区間とを前記時間軸を用いた画像表示により表示する機能を有する画像表示部を備えていることを特徴とする請求項15に記載の楽曲再生装置。

21. 前記複数の楽曲構造特徴区間は、楽曲のサビ部分を示す複数のサビ区間と、繰り返し部分を示す複数の繰り返し区間とからなり、

前記画像表示部は前記複数のサビ区間と前記複数の繰り返し区間とを区別して表示する機能を有していることを特徴とする請求項20に記載の楽曲再生装置。

22. 前記インターフェースは、前記複数の楽曲構造特徴区間から任意の楽

曲構造特徴区間を選択するために操作者が操作する 1 以上の選択ボタンを有している請求項 20 に記載の楽曲再生方法。

23. 前記 1 以上の選択ボタンには、前記複数の楽曲構造特徴区間から前記サビ区間だけを選択する 1 以上のサビ区間選択ボタンが含まれていることを特徴とする請求項 22 に記載の楽曲再生装置。

24. 前記 1 以上のサビ区間選択ボタンは前のサビ区間を選択する第 1 のサビ区間選択ボタンと、後のサビ区間を選択する第 2 のサビ区間選択ボタンとからなる請求項 22 に記載の楽曲再生装置。

25. 前記 1 以上の選択ボタンには、前記複数の繰り返し区間を選択する際に前の繰り返し区間を選択する第 1 の繰り返し区間選択ボタンと、後の繰り返し区間を選択する第 2 の繰り返し区間選択ボタンが含まれていることを特徴とする請求項 22 に記載の楽曲再生装置。

26. 前記 1 以上の選択ボタンは前記画像表示部に表示された画像表示ボタンである請求項 22 に記載の楽曲再生装置。

27. 前記複数の楽曲構造特徴区間は、楽曲のサビ部分を示す複数のサビ区間と、繰り返し区間を示す複数の繰り返し区間とからなり、

前記画像表示部は前記複数のサビ区間と前記複数の繰り返し区間とを区別して表示し、

前記画像表示部に示された前記複数のサビ区間及び前記複数の繰り返し区間の表示を、前記複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するために操作者が操作する前記 1 以上の選択ボタンとするように前記画像表示部が構成されていることを特徴とする請求項 26 に記載の楽曲再生装置。

28. 音楽音響データを再生する音楽音響データ再生装置を用いて、楽曲中の複数の楽曲構造特徴区間から選択した任意の楽曲構造特徴区間を再生する際に、前記楽曲構造特徴区間の選択と選択結果の伝送に用いられるインターフェースであって、

前記複数の楽曲構造特徴区間に含まれるサビ区間を選択するために操作者が操作するサビ区間選択ボタンを含む 1 以上の選択ボタンを有していることを特徴とするインターフェース。



29. 前記1以上の選択ボタンは前のサビ区間を選択する第1のサビ区間選択ボタンと、後のサビ区間を選択する第2のサビ区間選択ボタンとからなる請求項28に記載のインターフェース。

30. 前記複数の楽曲構造特徴区間に含まれる複数の繰り返し区間を選択する際に、前の繰り返し区間を選択する第1の繰り返し区間選択ボタンと、後の繰り返し区間を選択する第2の繰り返し区間選択ボタンをさらに備えていることを特徴とする請求項28に記載のインターフェース。

31. 前記音楽音響データの長さを時間軸を用いた画像表示により表示し、併せて前記音楽音響データ再生装置が前記音楽音響データを再生している前記音楽音響データの再生位置と前記複数の楽曲構造特徴区間とを前記時間軸を用いた画像表示により表示する画像表示部を更に有していることを特徴とする請求項28に記載のインターフェース。

32. 前記複数の楽曲構造特徴区間は、楽曲のサビ部分を示す複数のサビ区間と、繰り返し区間を示す複数の繰り返し区間とからなり、

前記画像表示部は前記複数のサビ区間と前記複数の繰り返し区間とを区別して表示し、

前記画像表示部に示された前記複数のサビ区間及び前記複数の繰り返し区間の表示を、前記複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するために操作者が操作する前記1以上の選択ボタンとするように構成されていることを特徴とする請求項31に記載のインターフェース。

33. コンピュータと、前記コンピュータからの指令に基づいて動作するディスプレイと、前記コンピュータからの指令で音楽音響データを再生する音楽音響データ再生装置とから構成された装置を用いて、楽曲中の複数の楽曲構造特徴区間から選択した任意の楽曲構造特徴区間を再生する機能を前記コンピュータに実現させるためのプログラムであって、

前記複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するインターフェースを前記ディスプレイ上に構成する機能と、

再生する前記音楽音響データに対応して予め定められた前記複数の楽曲構造特徴区間を指定するための楽曲構造特徴区間指定データを格納する楽曲構造特徴区

間指定データ格納手段を構成する機能と、

前記インターフェースによって選択された前記任意の楽曲構造特徴区間を指定する前記楽曲構造特徴区間指定データを前記音楽音響データ再生装置に与える指定データ付与手段を構成する機能と、

前記指定データ付与手段から付与された前記前記楽曲構造特徴区間指定データによって指定された前記楽曲構造特徴区間を再生するように前記音楽音響データ再生装置に指令を与える機能とを前記コンピュータに実現させる構成を備えたことを特徴とする楽曲再生装置実現用プログラム。

34. 前記インターフェースを前記ディスプレイ上に構成する機能は、前記複数の楽曲構造特徴区間に含まれるサビ区間を選択するために操作者が操作するサビ区間選択ボタンを含む1以上の選択ボタンを前記ディスプレイ上に形成する機能を有していることを特徴とする請求項33に記載の楽曲再生装置実現用プログラム。

35. 前記インターフェースを前記ディスプレイ上に構成する機能は、前記音楽音響データの長さを時間軸を用いた画像表示により前記ディスプレイ上に表示し、併せて前記音楽音響データ再生装置が前記音楽音響データを再生している前記音楽音響データの再生位置と前記複数の楽曲構造特徴区間とを前記時間軸を用いた画像表示により前記ディスプレイ上に表示する機能を更に有していることを特徴とする請求項33に記載の楽曲再生装置実現用プログラム。

36. 前記複数の楽曲構造特徴区間は、楽曲のサビ部分を示す複数のサビ区間と、繰り返し区間を示す複数の繰り返し区間とからなり、

前記インターフェースを前記ディスプレイ上に構成する機能は、前記複数のサビ区間と前記複数の繰り返し区間とを区別して前記ディスプレイ上に表示し、前記ディスプレイ上に示された前記複数のサビ区間及び前記複数の繰り返し区間の表示を、前記複数の楽曲構造特徴区間から任意の楽曲構造特徴区間を選択するために操作者が操作する前記1以上の選択ボタンとする機能をさらに有していることを特徴とする請求項33に記載の楽曲再生装置実現用プログラム。

37. 音楽音響データ中から楽曲構造特徴区間を抽出する方法であって、対象とする前記音楽音響データを複数の試聴者が再生した際の再生行動の統計

データに基づいて前記楽曲構造特徴区間を抽出することを特徴とする楽曲構造特徴区間の抽出方法。

38. 複数の試聴者による再生回数が多い区間を前記複数の楽曲構造特徴区間と定めることを特徴とする請求項37に記載の楽曲構造特徴区間の抽出方法。

39. ネットワークに接続された複数台の試聴用の楽曲再生装置から前記複数の試聴者による再生回数が多い区間のデータを集めることを特徴とする請求項38に記載の楽曲構造特徴区間の抽出方法。

40. 前記データをヒストグラムとして視覚により認識可能な状態とし、前記ヒストグラムから前記複数の楽曲構造特徴区間を特定することを特徴とする請求項39に記載の楽曲構造特徴区間の抽出方法。

41. ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出する方法であって、

前記音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出ステップと、

前記音楽音響データについて求めた複数の前記音響特徴量の相互間の類似度を求める類似度演算ステップと、

前記類似度に基づいて前記音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップステップと、

リストアップされた前記複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある1以上の前記繰り返し区間を時間軸上に統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定ステップと、

前記複数種類の統合繰り返し区間列から前記サビ区間を決定するサビ区間決定ステップとからなることを特徴とする音楽音響データ中のサビ区間を検出する方法。

42. 前記特徴量抽出ステップで求める音響特徴量は、1オクターブの範囲に含まれる12の音名の周波数のパワーを複数のオクターブに渡ってそれぞれ加算して得た12次元クロマベクトルである請求項41に記載の音楽音響データ中のサビ区間を検出する方法。

4 3. 前記類似度演算ステップでは、今回求めた前記音響特徴量と先に求めた全ての前記音響特徴量との間の前記類似度を求めることを特徴とする請求項 4 2 に記載の音楽音響データ中のサビ区間を検出する方法。

4 4. 前記類似度演算ステップでは、時刻  $t$  の前記 1 2 次元クロマベクトルとそれよりラグ  $l$  ( $0 \leq l \leq t$ ) だけ過去の全ての前記 1 2 次元クロマベクトルとの前記類似度を求め、

前記繰り返し区間リストアップステップでは、一方の軸を時間軸とし他方の軸をラグ軸とし、予め定めた時間長さ以上前記類似度が予め定めた閾値以上ある場合には前記類似度が前記予め定めた閾値以上である部分の長さに対応する時間長さを有する類似線分を前記時間軸を基準にした前記繰り返し区間としてリストアップすることを特徴とする請求項 4 3 に記載の音楽音響データ中のサビ区間を検出する方法。

4 5. 統合繰り返し区間決定ステップでは、前記時間軸の共通区間に存在するリストアップした前記類似線分どうしをそれぞれグルーピングにより統合して前記統合繰り返し区間と定め、

複数の前記統合繰り返し区間を、前記共通区間の前記時間軸上の存在位置及び長さでグルーピングされる前記類似線分の前記ラグ軸で見た位置関係とに基づいて前記複数種類の前記統合繰り返し区間列に分類することを特徴とする請求項 4 4 に記載の音楽音響データ中のサビ区間を検出する方法。

4 6. 統合繰り返し区間決定ステップでは、前記統合繰り返し区間に含まれない最初の繰り返し区間を補足して前記統合繰り返し区間列を作成する請求項 4 5 に記載の音楽音響データ中のサビ区間を検出する方法。

4 7. 前記楽曲は転調を含んでおり、

前記特徴量抽出ステップでは、前記 1 2 次元クロマベクトルからなる前記音響特徴量を 1 転調幅ずつ 1 1 転調幅までシフトして得た転調幅の異なる 1 2 種類の前記音響特徴量を求め、

前記類似度演算ステップでは、今回求めた前記音響特徴量と先に求めた全ての 1 2 種類の前記音響特徴量との間の前記類似度を、時刻  $t$  の今回の前記音響特徴量を表す前記クロマベクトルとそれよりラグ  $l$  ( $0 \leq l \leq t$ ) だけ過去の全ての

1 2 種類の前記音響特徴量を表す前記クロマベクトルとの間の類似度として演算し、

前記繰り返し区間リストアップステップでは、1 2 種類の前記音響特徴量ごとに、一方の軸を時間軸  $t$  とし他方の軸をラグ  $l$  とし、予め定めた時間長さ以上前記類似度が予め定めた閾値以上である部分の長さに対応する時間長さを有する類似線分を前記時間軸を基準にした前記繰り返し区間としてそれぞれ 1 2 種類のリストをリストアップすることを特徴とする請求項 4 1 に記載の音楽音響データ中のサビ区間を検出する方法。

4 8. 統合繰り返し区間決定ステップでは、前記 1 2 種類のリストごとに、前記時間軸の共通区間に存在するリストアップした前記類似線分どうしをそれぞれグルーピングにより統合して統合繰り返し区間と定め、

さらに前記 1 2 種類のリストについて定めた複数の前記統合繰り返し区間を、前記共通区間の前記時間軸上の存在位置及び長さ、グルーピングされる前記類似線分の前記ラグ軸で見た位置関係とに基づいて前記複数種類の転調を考慮した前記複数種類の統合繰り返し区間列に分類化することを特徴とする請求項 4 7 に記載の音楽音響データ中のサビ区間を検出する方法。

4 9. 前記サビ区間決定ステップでは、前記統合繰り返し区間列に含まれる前記統合繰り返し区間の前記類似度の平均と、数と長さに基づいて該統合繰り返し区間列に含まれる前記統合繰り返し区間のサビらしさを求め、最もサビらしさの高い前記統合繰り返し区間列に含まれる前記統合繰り返し区間を前記サビ区間として決定することを特徴とする請求項 4 1 に記載の音楽音響データ中のサビ区間を検出する方法。

5 0. ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出して表示手段に表示する装置であって、

前記音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出手段と、

前記音楽音響データについて求めた複数の前記音響特徴量の相互間の類似度を求める類似度演算手段と、

前記類似度に基づいて前記音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップ手段と、

リストアップされた前記複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある1以上の前記繰り返し区間を統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定手段と、

前記複数種類の統合繰り返し区間列から前記サビ区間を決定するサビ区間決定手段とを具備し、

前記複数種類の統合繰り返し区間列が前記表示手段に表示され、

前記サビ区間を含む前記統合繰り返し区間列が他の前記統合繰り返し区間列とは異なる表示態様で表示されることを特徴とする音楽音響データ中のサビ区間を検出する装置。

51. ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出して表示手段に表示する装置であって、

前記音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出手段と、

前記音楽音響データについて求めた複数の前記音響特徴量の相互間の類似度を求める類似度演算手段と、

前記類似度に基づいて前記音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップ手段と、

リストアップされた前記複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある1以上の前記繰り返し区間を統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定手段と、

前記複数種類の統合繰り返し区間列から前記サビ区間を決定するサビ区間決定手段とを具備することを特徴とする音楽音響データ中のサビ区間を検出する装置

52. 統合繰り返し区間決定手段は、前記統合繰り返し区間に含まれない最

初の繰り返し区間を補足して前記統合繰り返し区間列を作成するように構成されている請求項 5 1 に記載の音楽音響データ中のサビ区間を検出する装置。

5 3. ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出して前記サビ区間を再生手段により再生する装置であって、

前記音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出手段と、

前記音楽音響データについて求めた複数の前記音響特徴量の相互間の類似度を求める類似度演算手段と、

前記類似度に基づいて前記音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップ手段と、

リストアップされた前記複数の繰り返し区間の相互関係を調べ、時間軸上の共通区間にある 1 以上の前記繰り返し区間を統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定手段と、

前記複数種類の統合繰り返し区間列から前記サビ区間を決定するサビ区間決定手段とを具備し、

前記複数種類の統合繰り返し区間列が選択的に前記再生手段で再生されることを特徴とする音楽音響データ中のサビ区間を検出する装置。

5 4. ある楽曲中で繰り返されるサビ区間を検出するためにその楽曲の音楽音響データ中からサビ区間に対応する部分を検出する方法をコンピュータを用いて実現するために用いられるプログラムであって、

前記音楽音響データから所定の時間単位で音響特徴量を順次求める特徴量抽出ステップと、

前記音楽音響データについて求めた複数の前記音響特徴量の相互間の類似度を求める類似度演算ステップと、

前記類似度に基づいて前記音楽音響データ中に繰り返し現れる複数の繰り返し区間をリストアップする繰り返し区間リストアップステップと、

リストアップされた前記複数の繰り返し区間の相互関係を調べ、時間軸上の共

通区間にある 1 以上の前記繰り返し区間を統合して一つの統合繰り返し区間を決定し、決定した複数の前記統合繰り返し区間を複数種類の統合繰り返し区間列に分類化する統合繰り返し区間決定ステップと、

前記複数種類の統合繰り返し区間列から前記サビ区間を決定するサビ区間決定ステップとを前記コンピュータに実行させるように構成されていることを特徴とするプログラム。

55. 統合繰り返し区間決定ステップでは、前記統合繰り返し区間に含まれない最初の繰り返し区間を補足して前記統合繰り返し区間列を作成する請求項 54 に記載のプログラム



図1

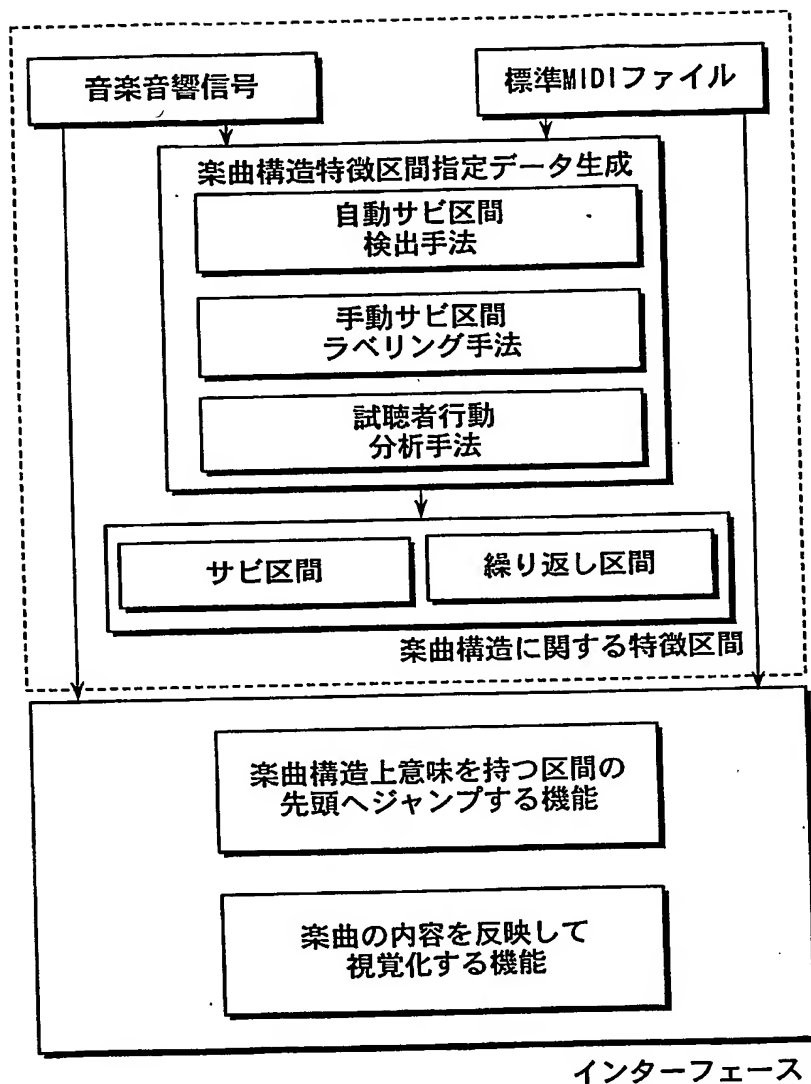


図2

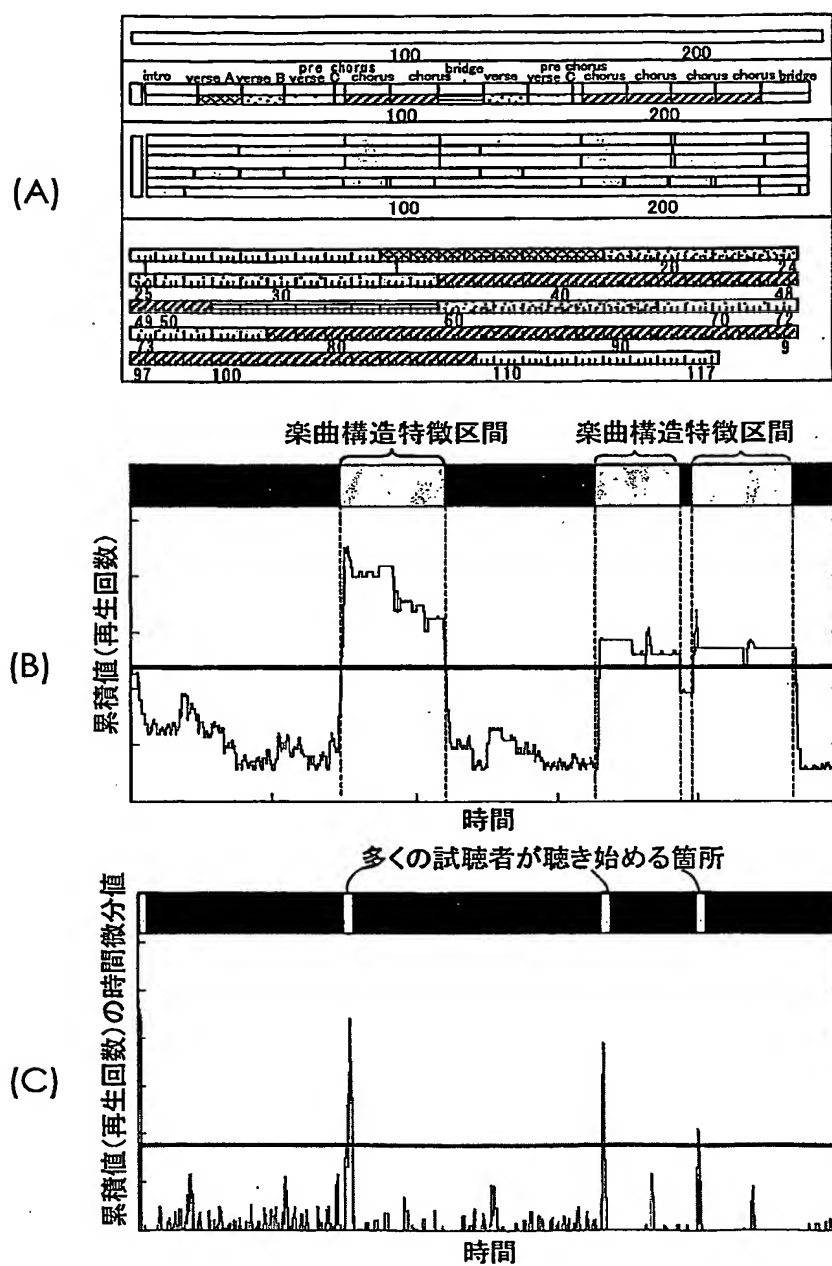


図3

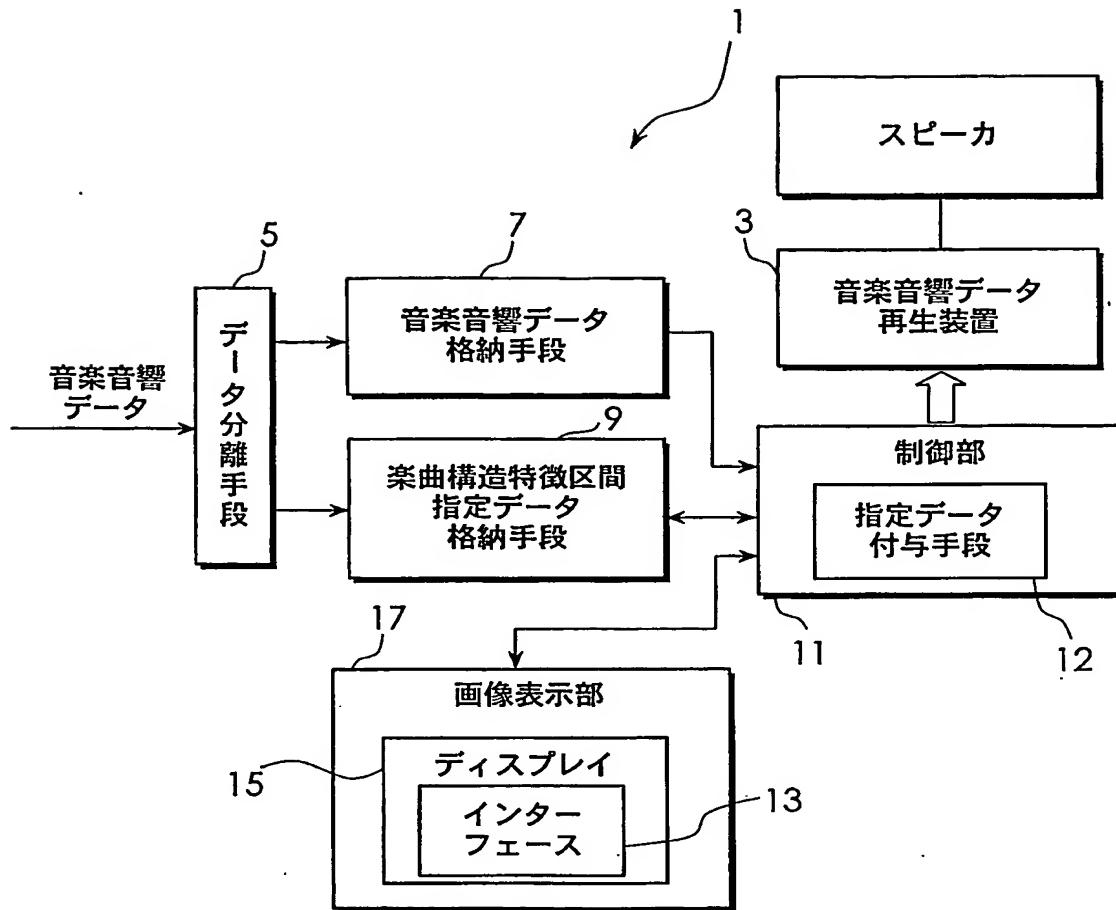


図4

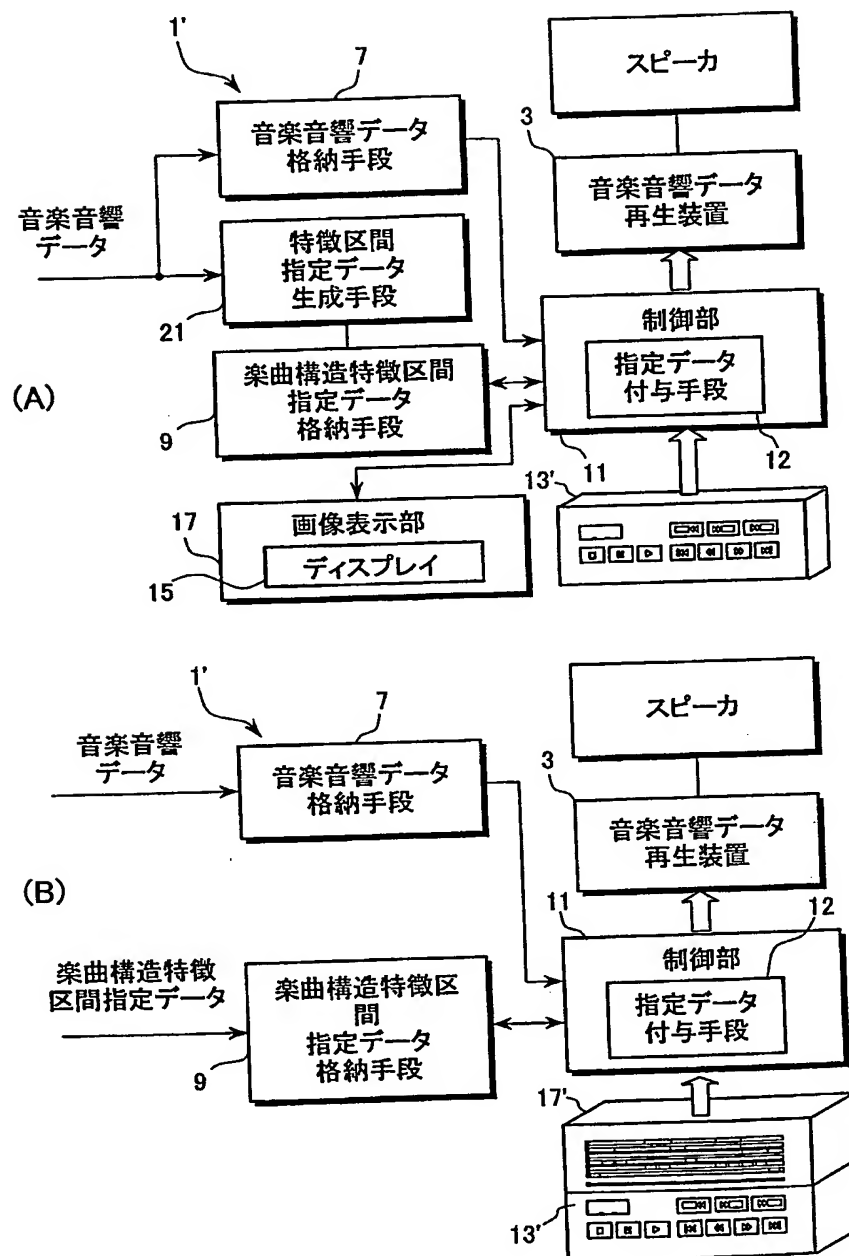


図5

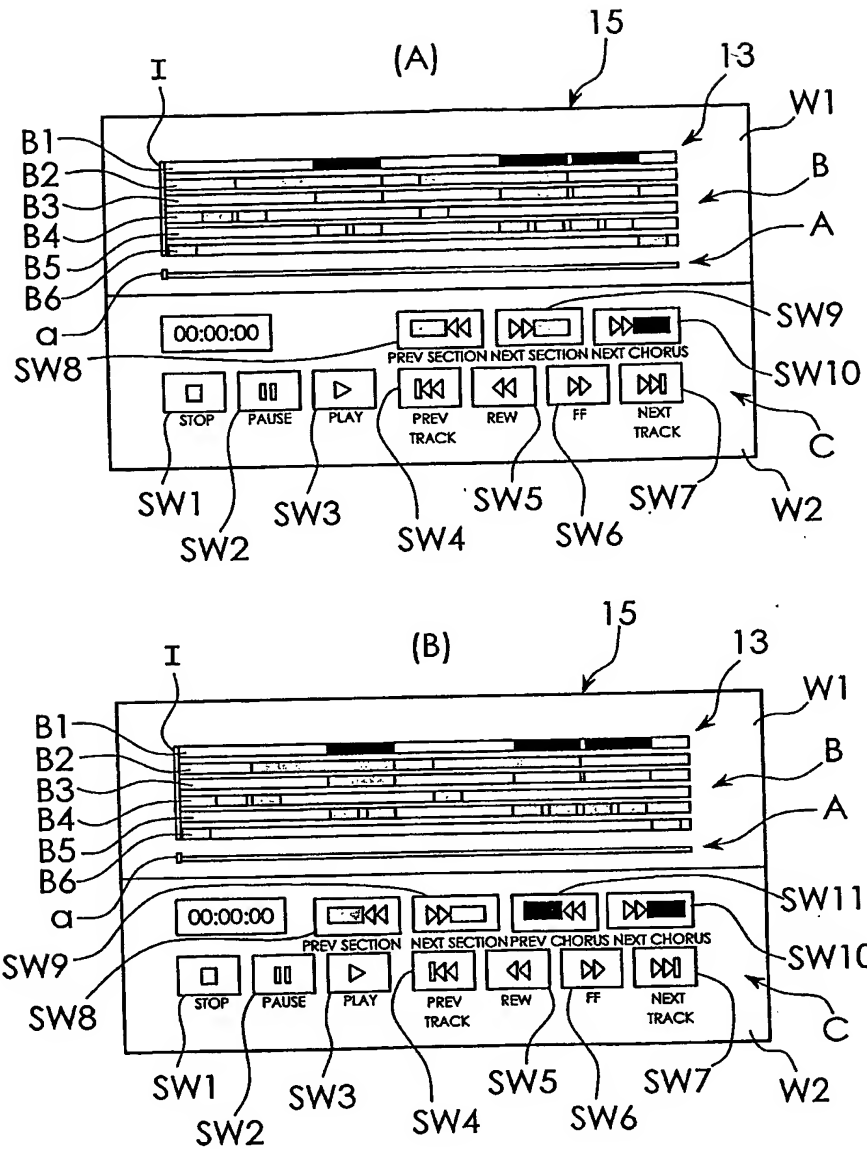


図6

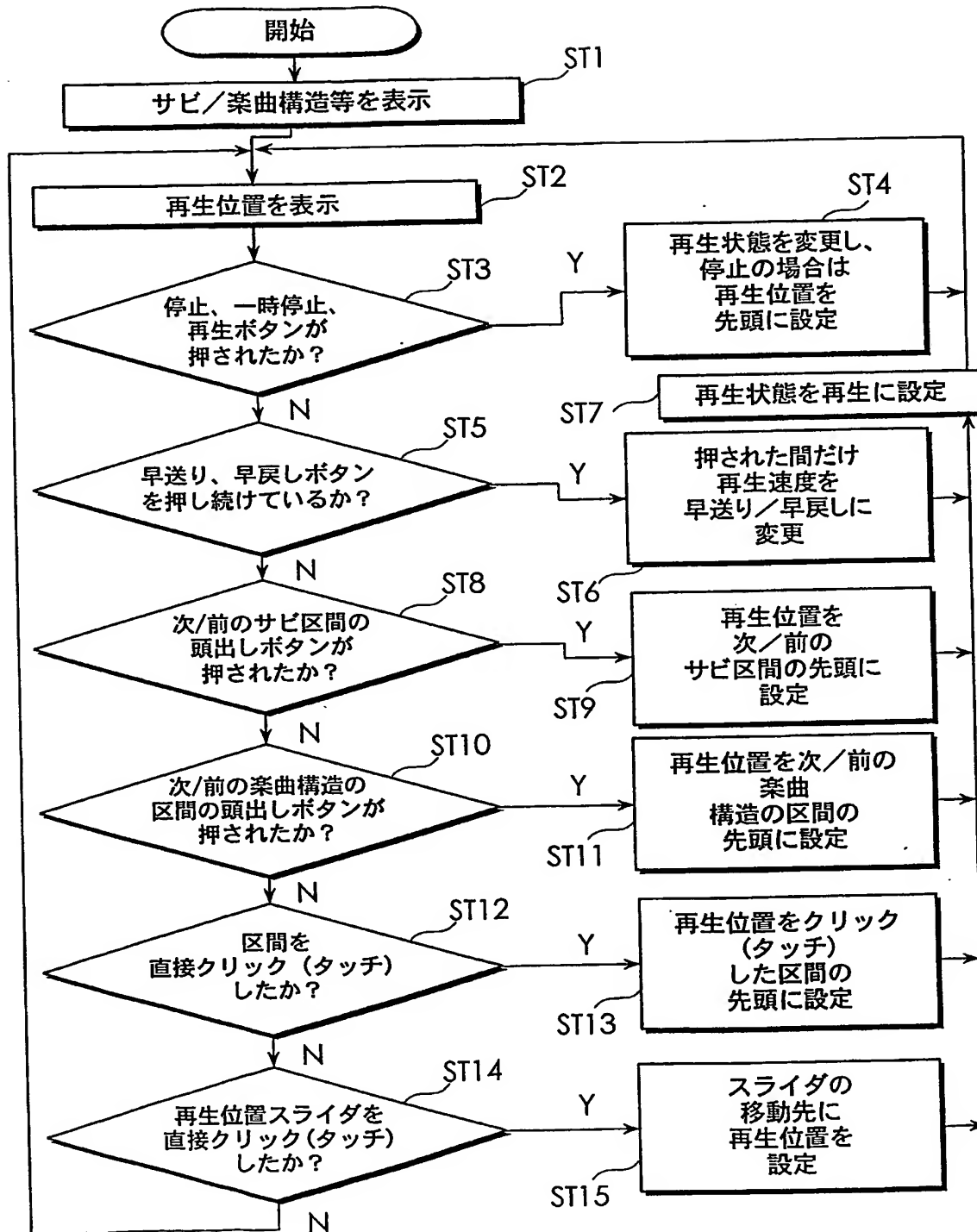


図7

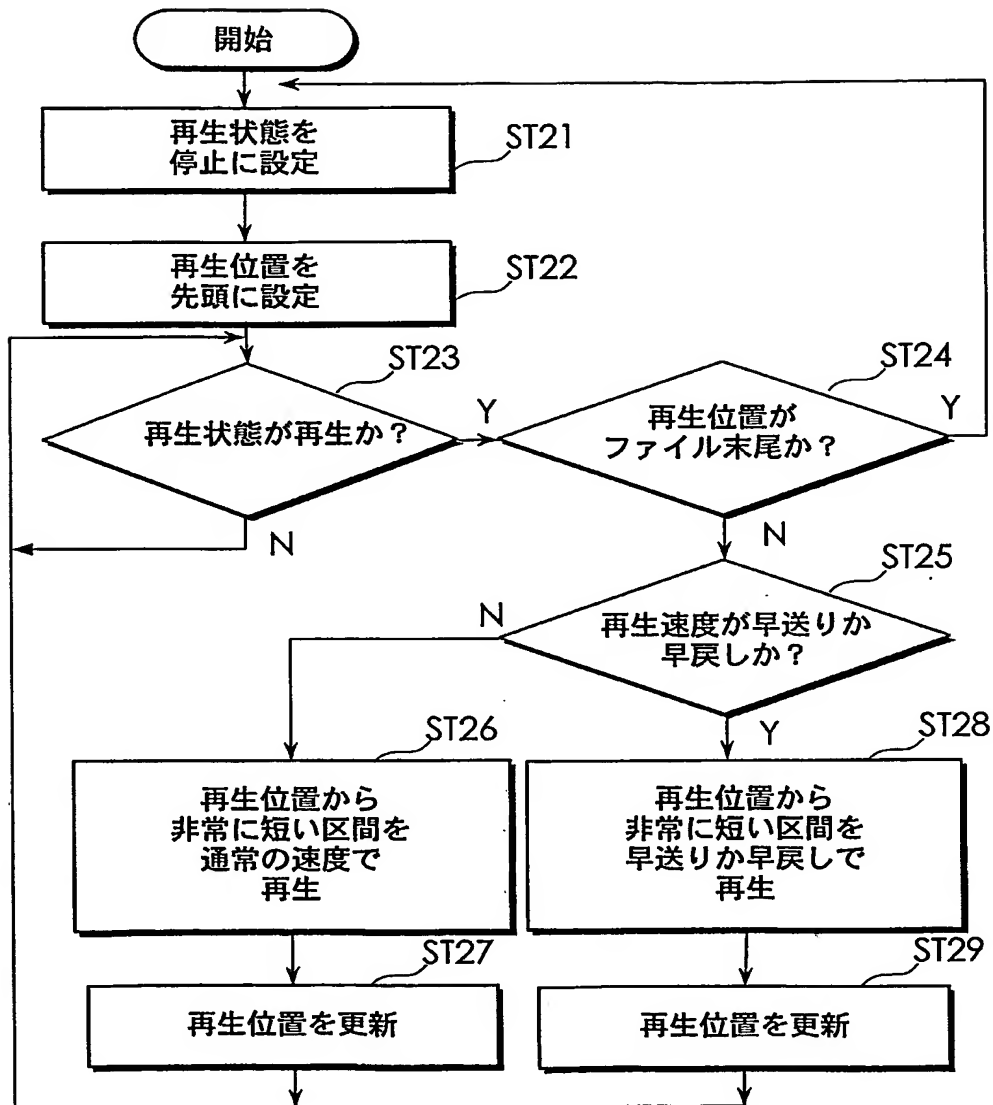


図 8

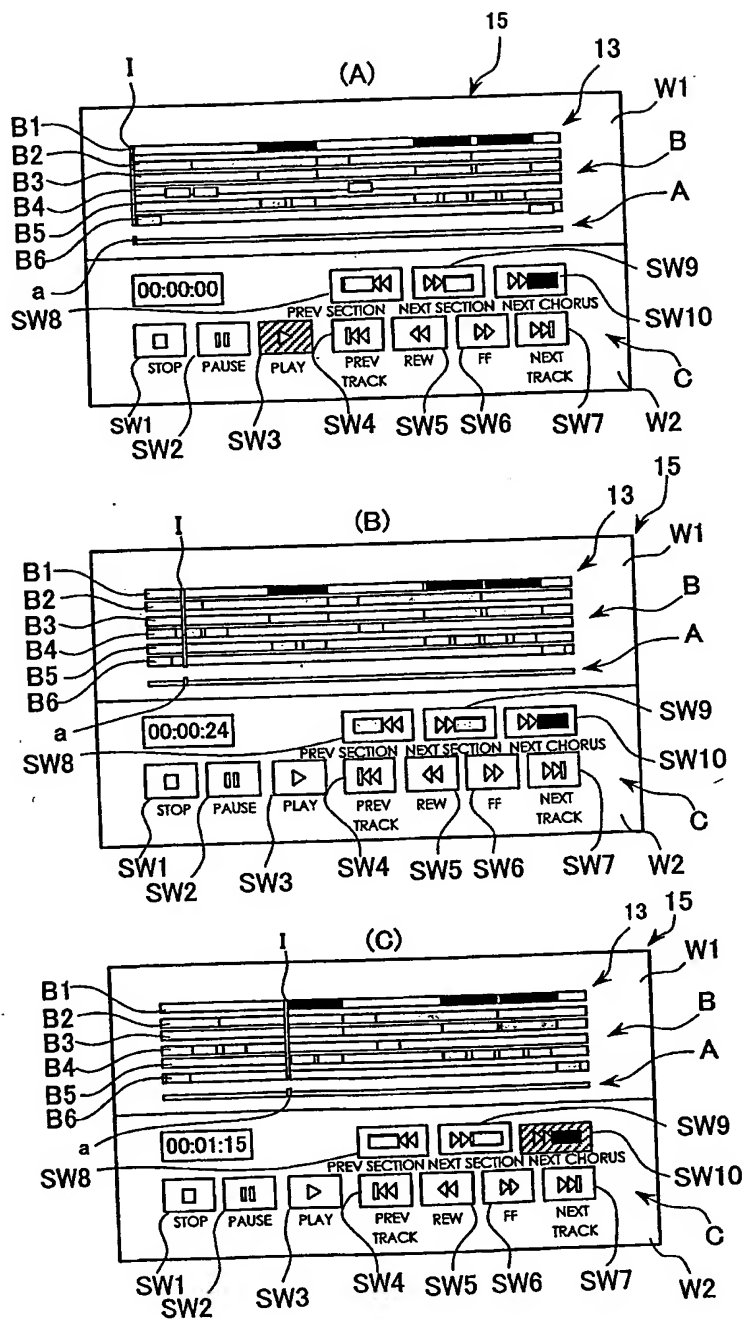




図9

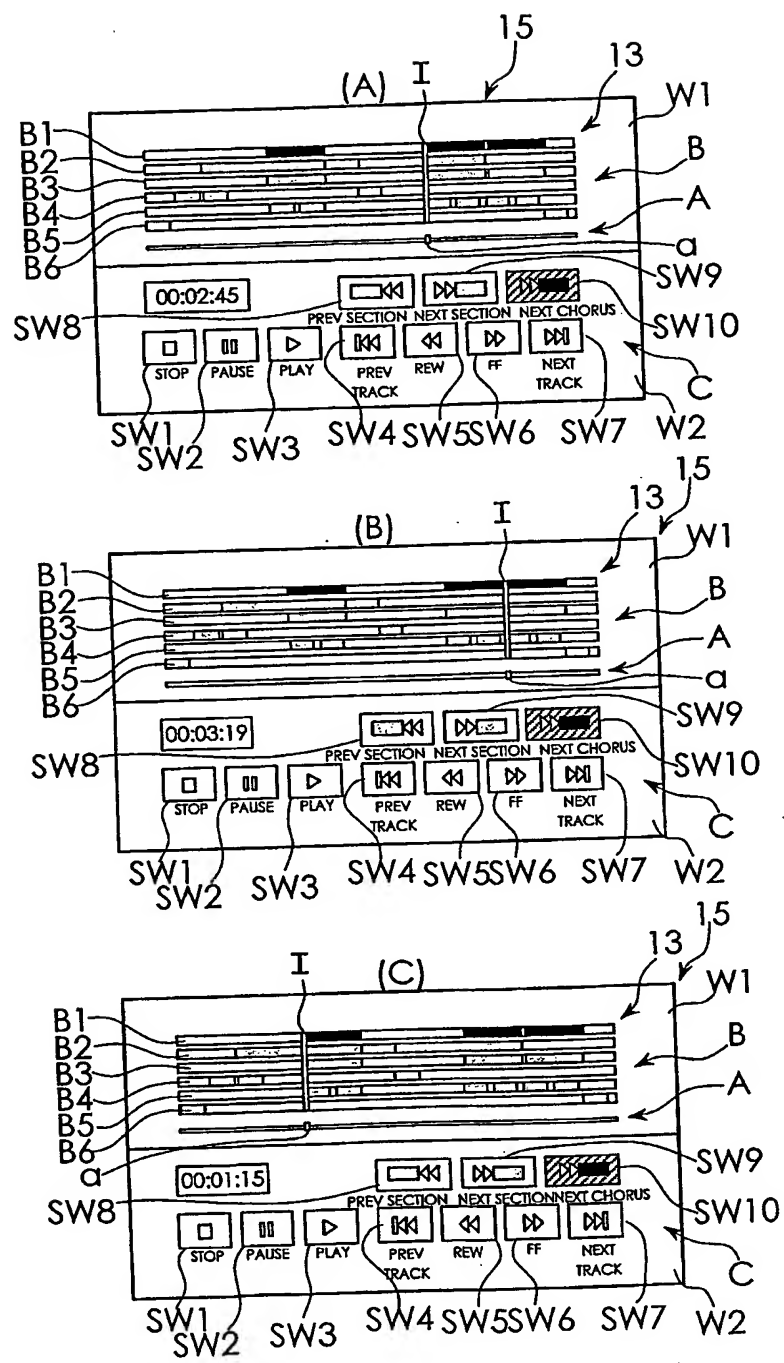


図 10

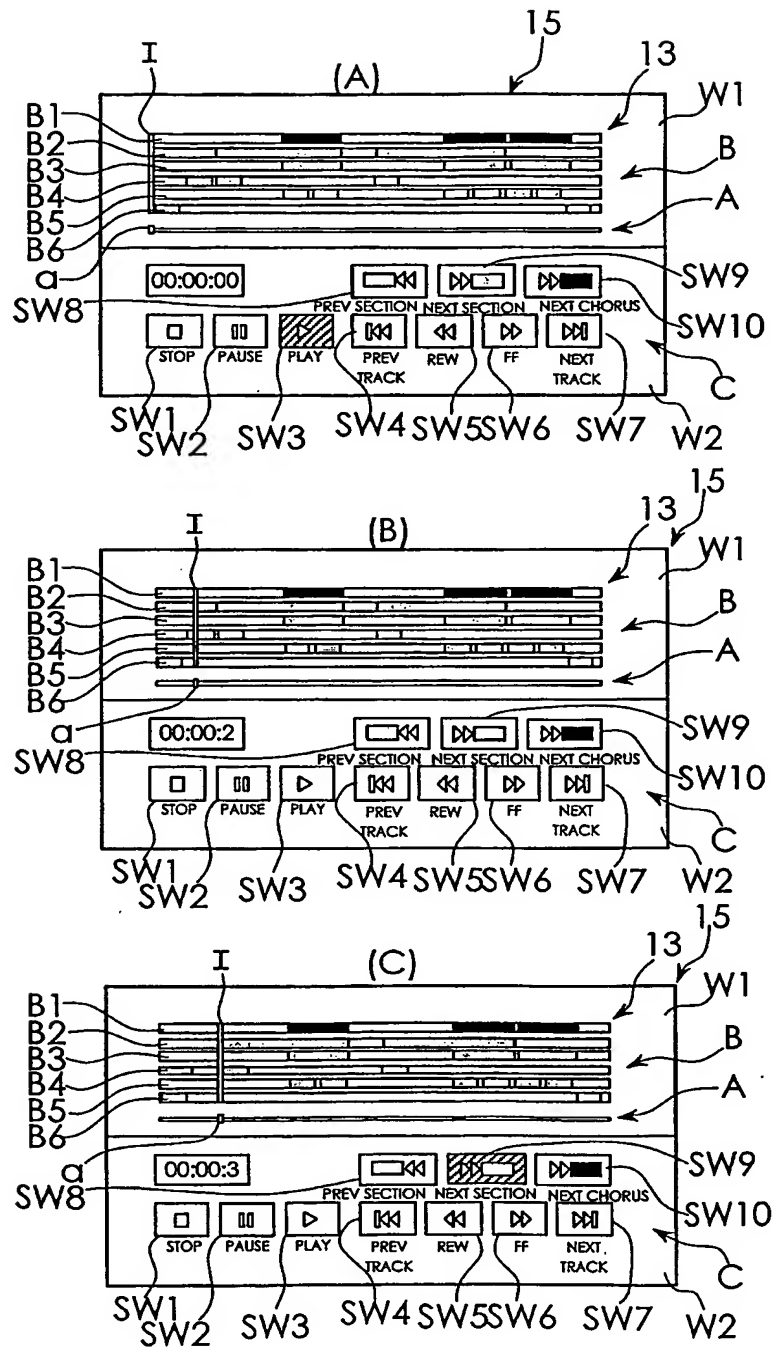


図 11

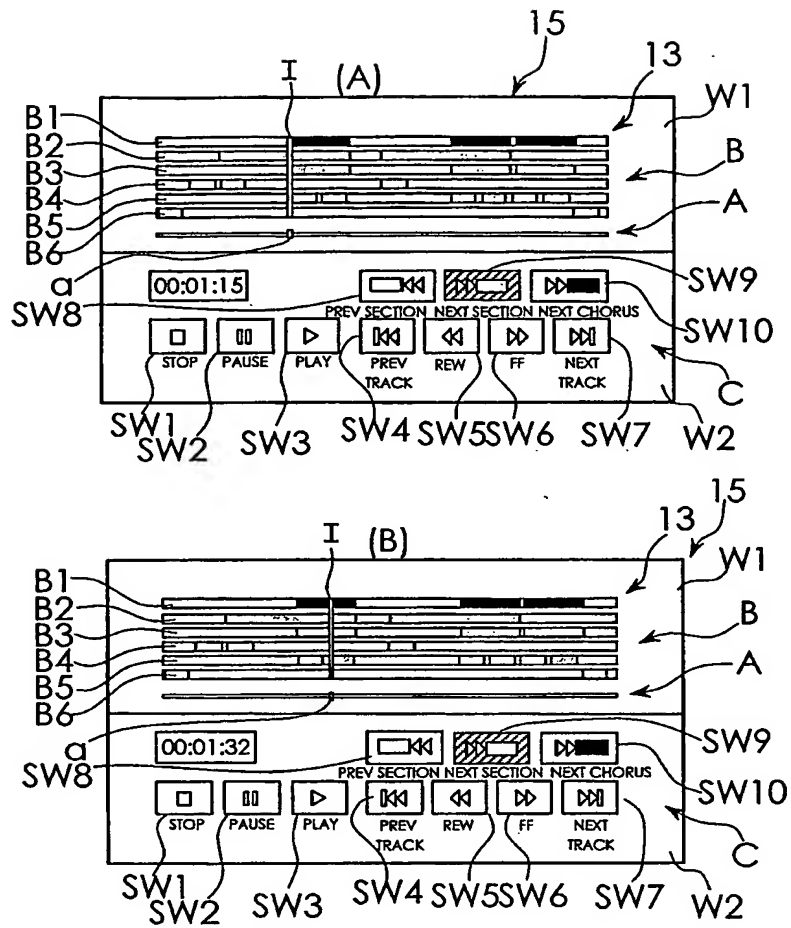


図12

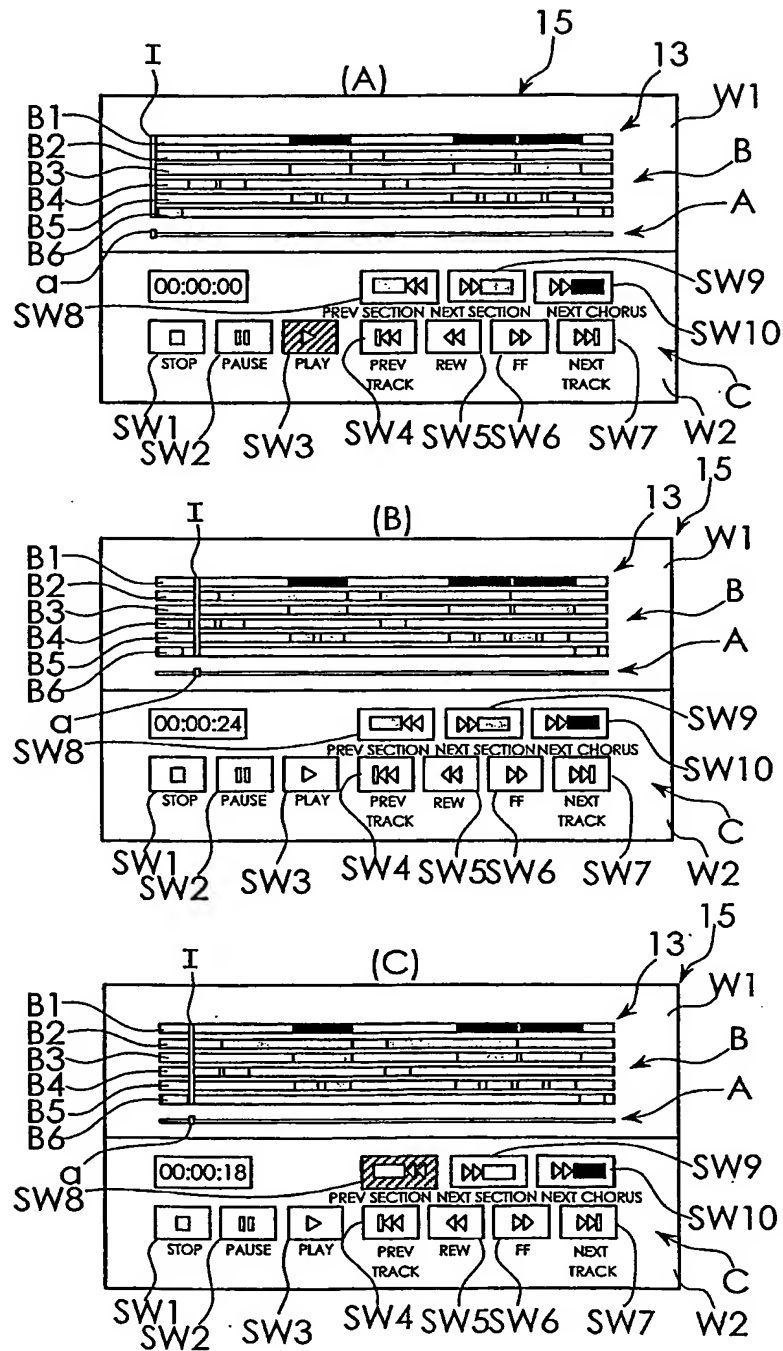


图13

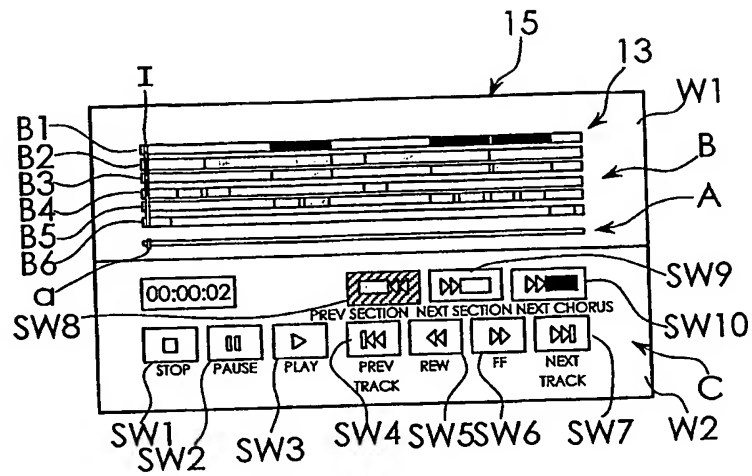


図14

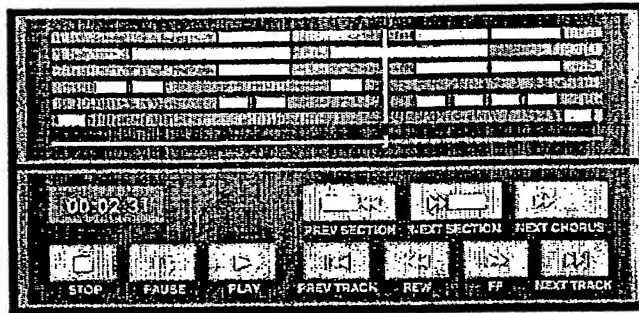


図15

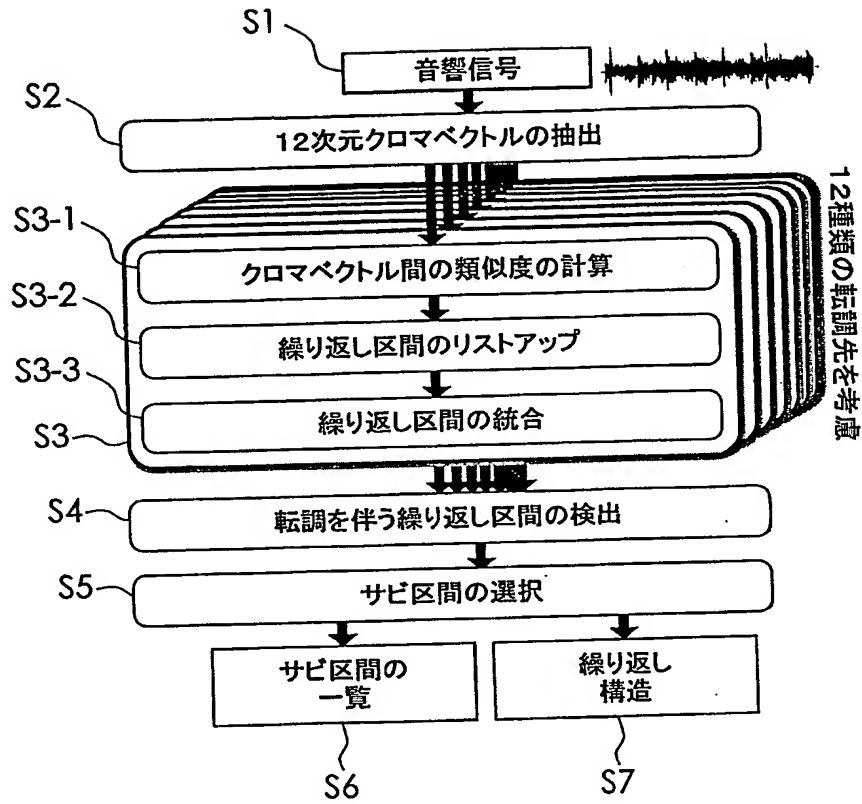


図16

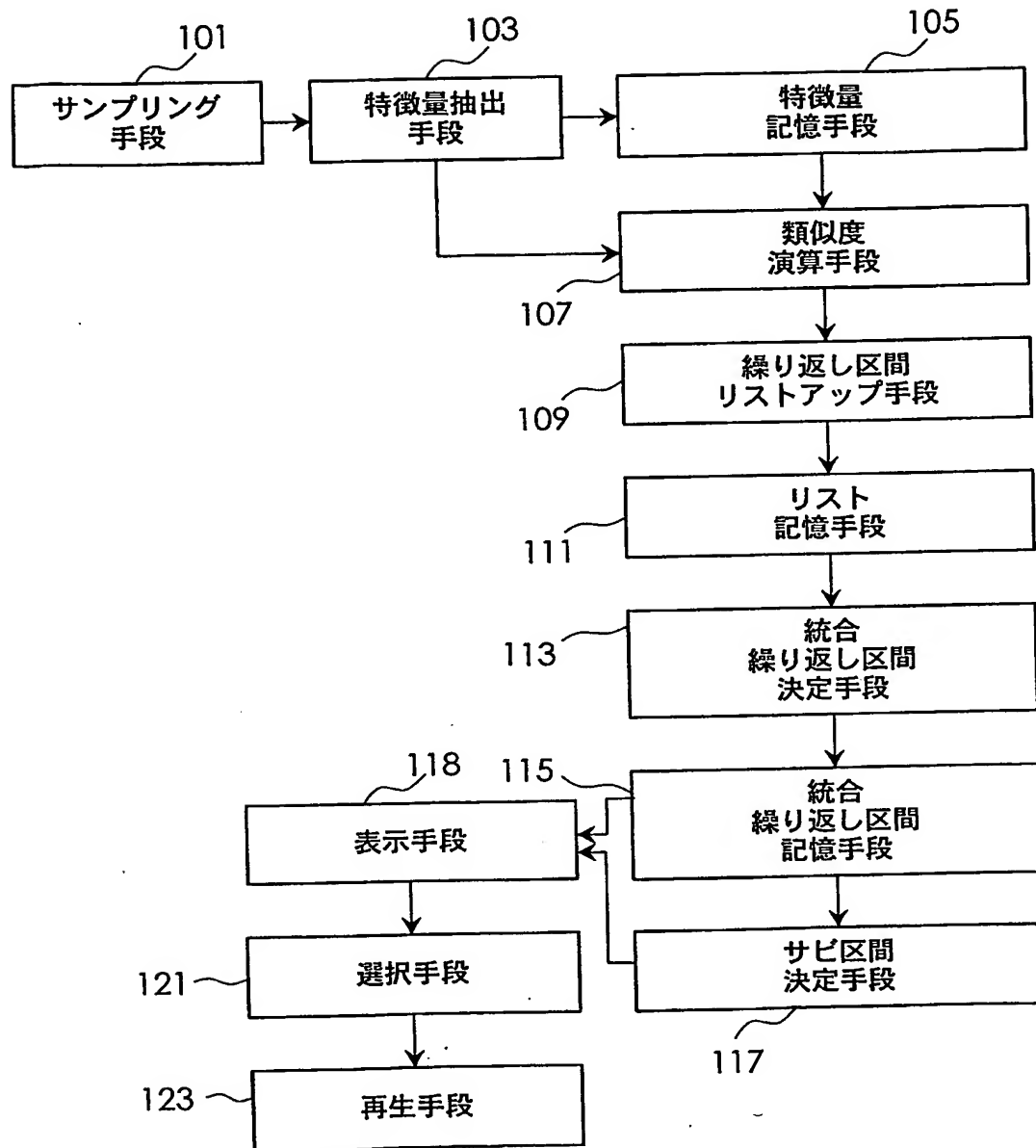




図17

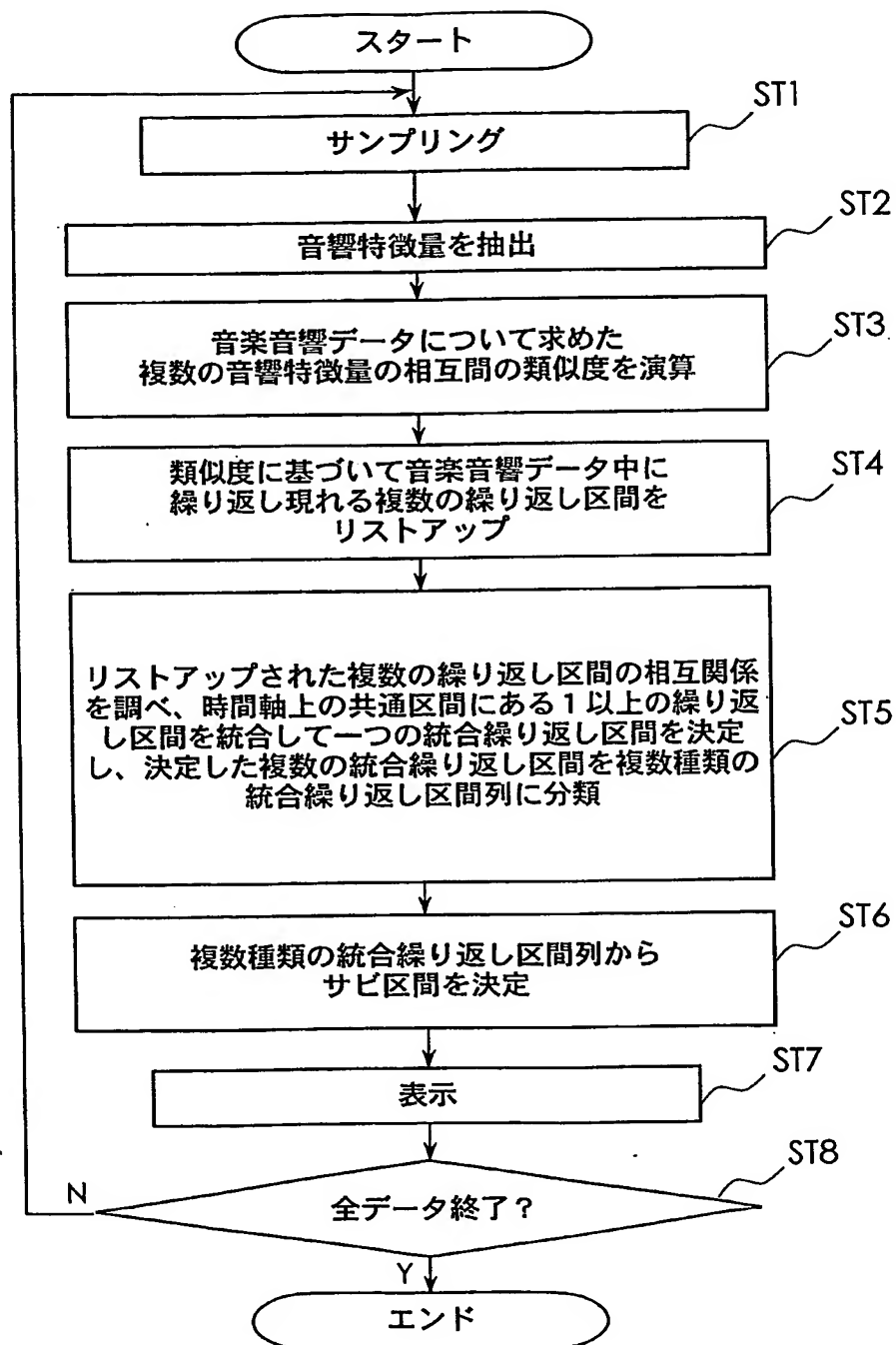


図18

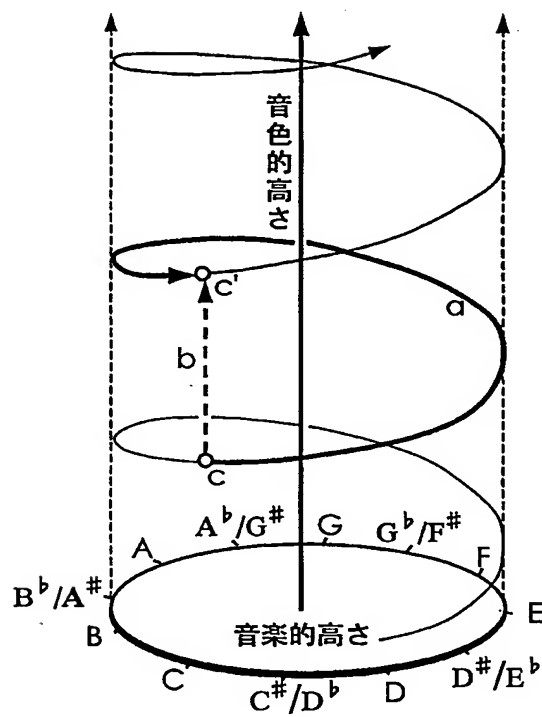


図19

## STFTパワースペクトル

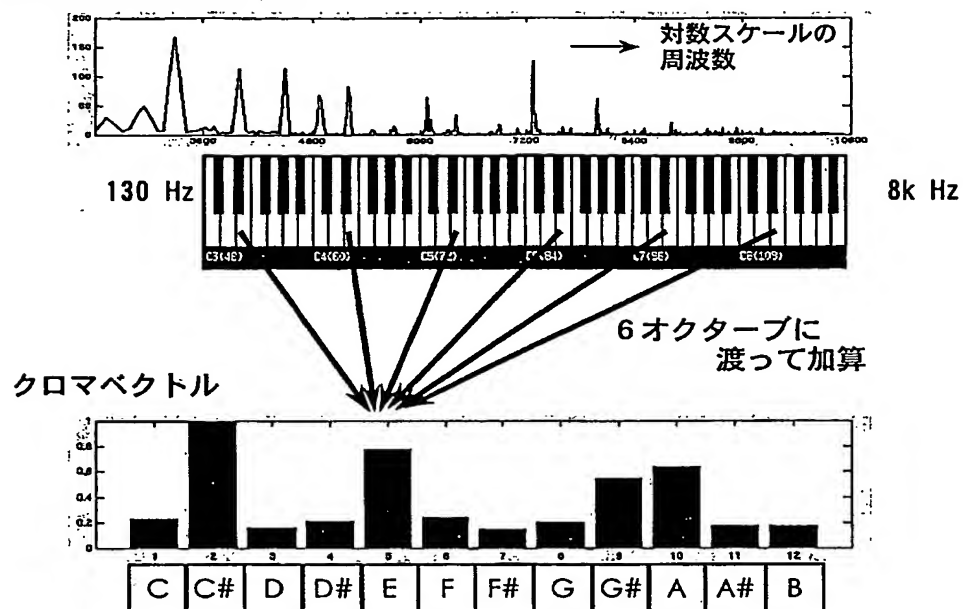


図20

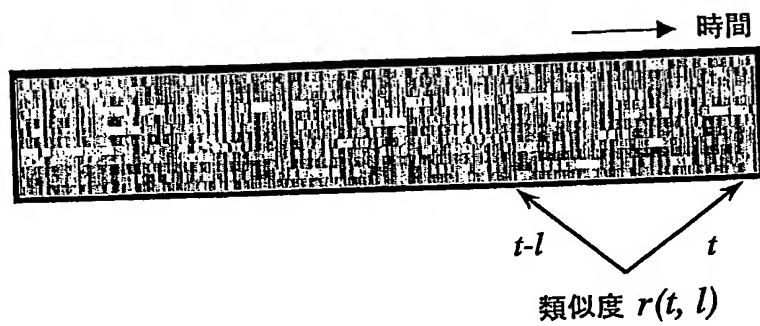


図21

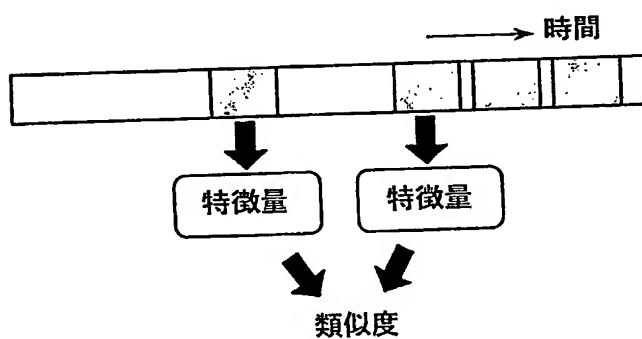


図22

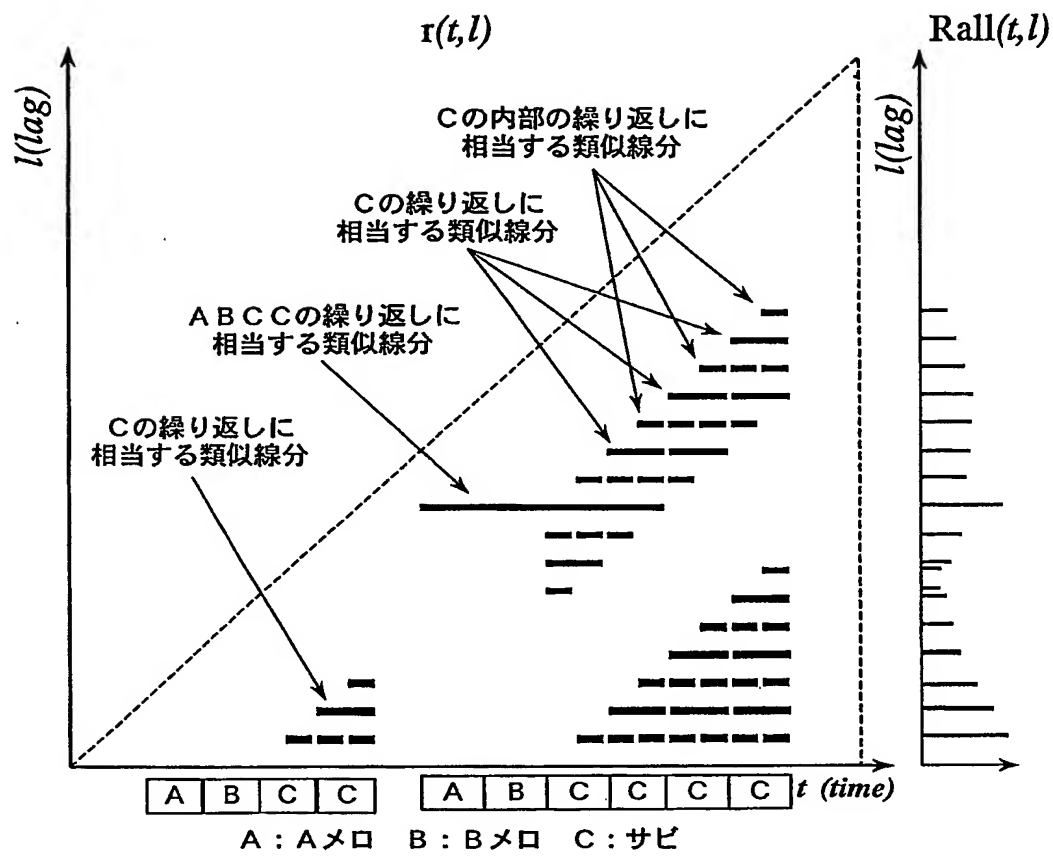


図23

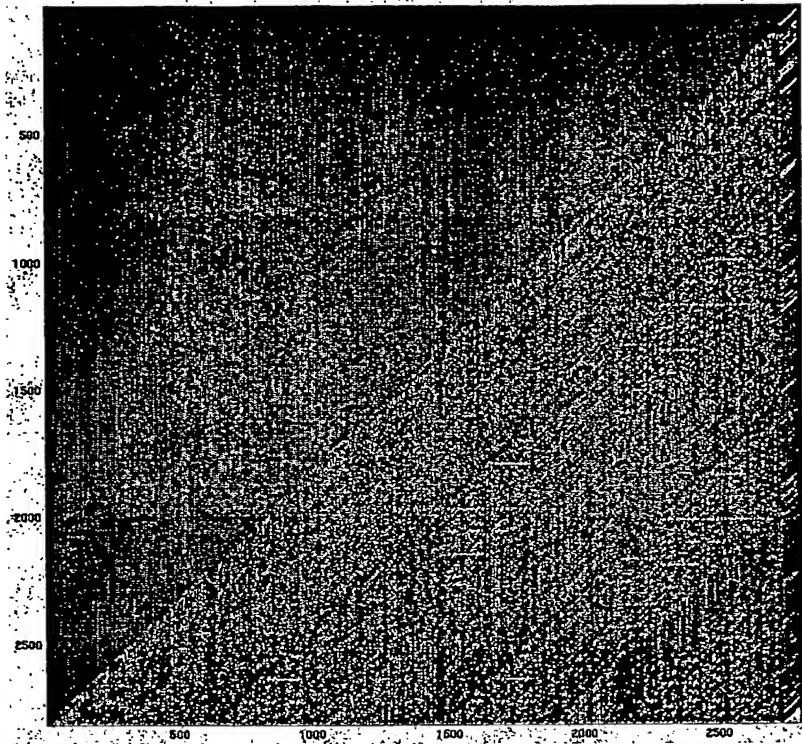


図24

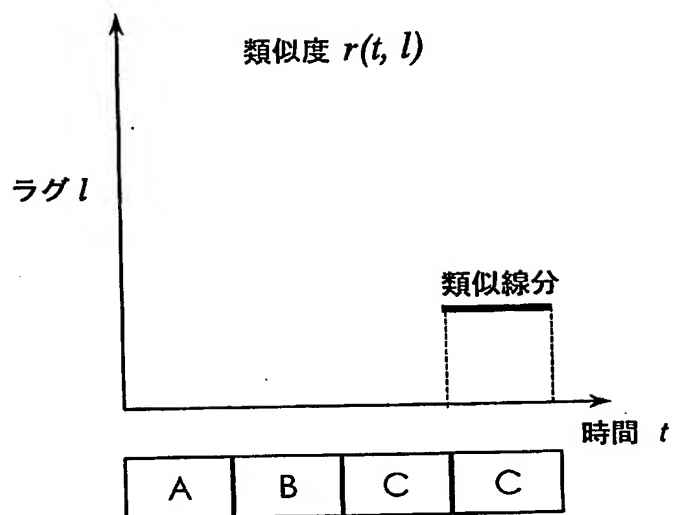


図25

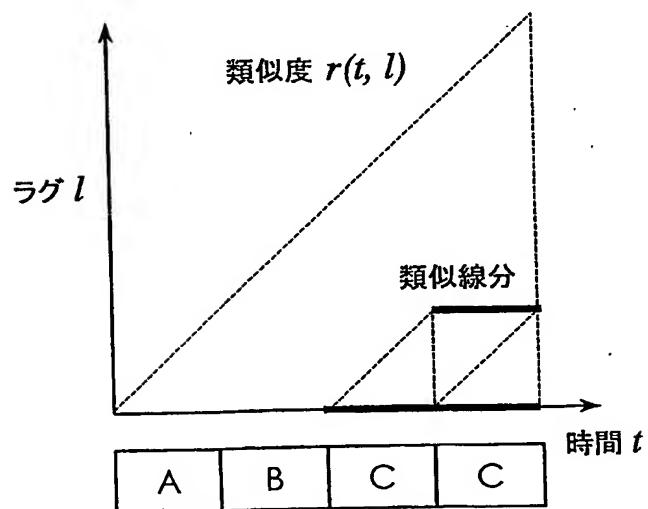


図26

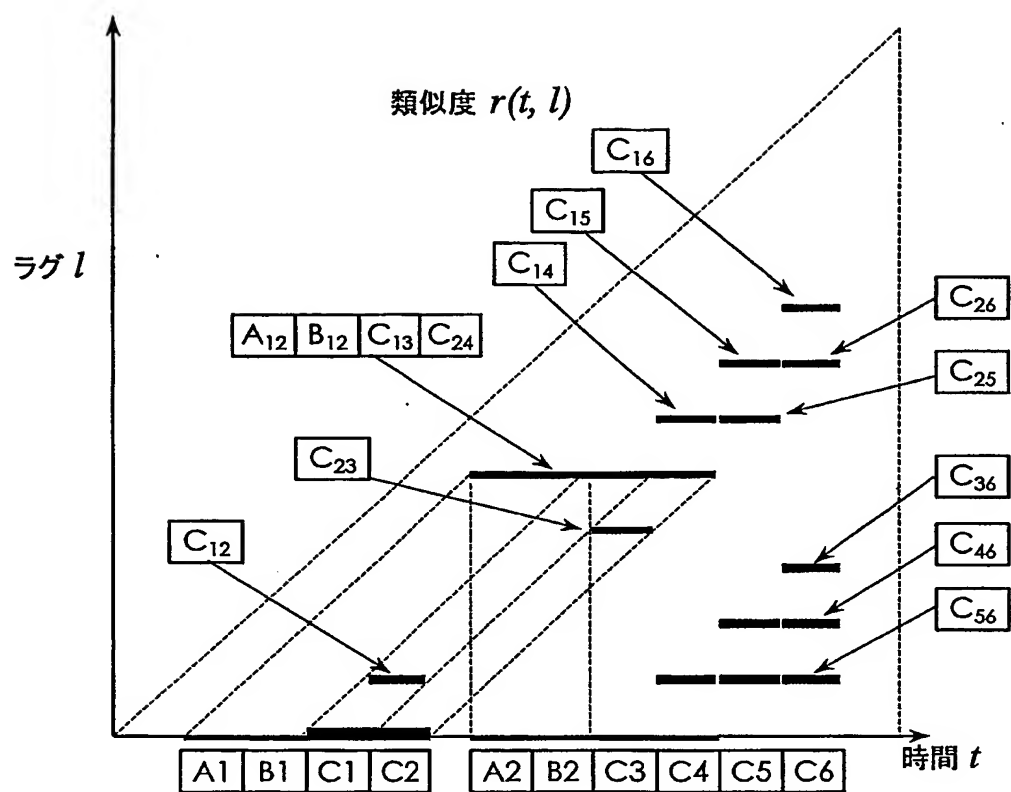




図27

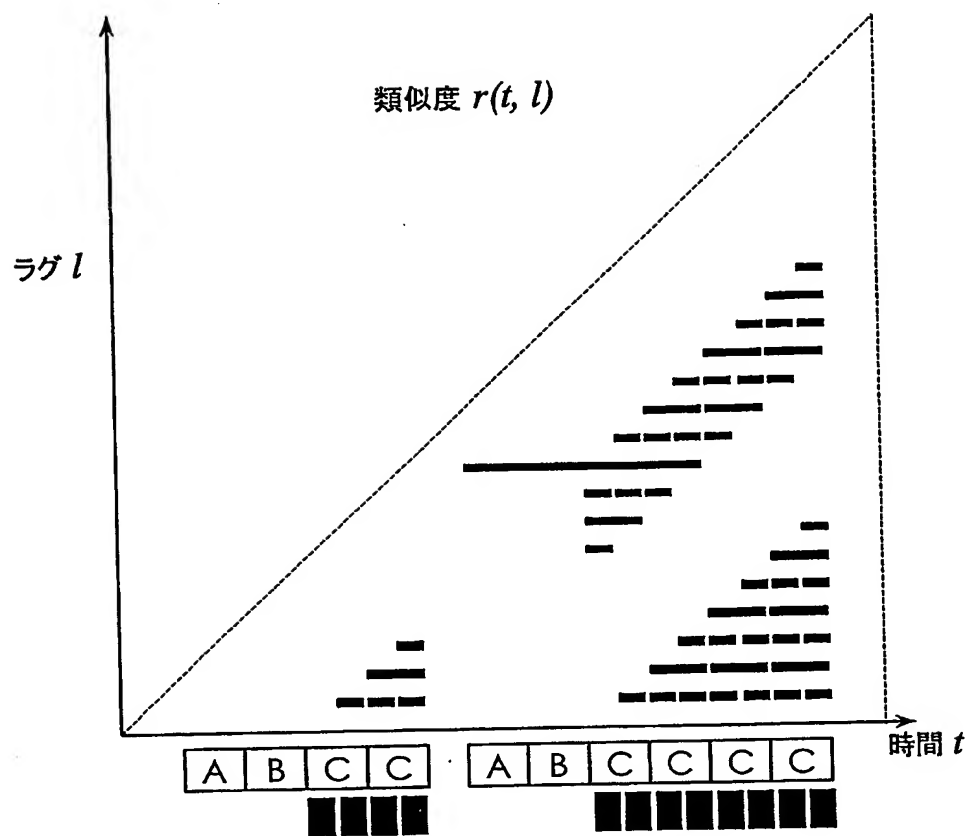


図28

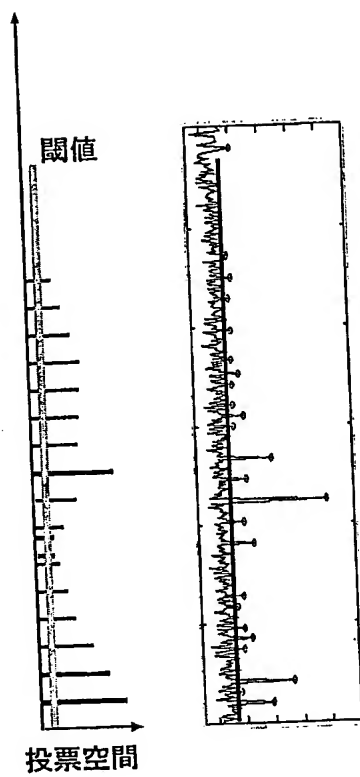


図29



図30

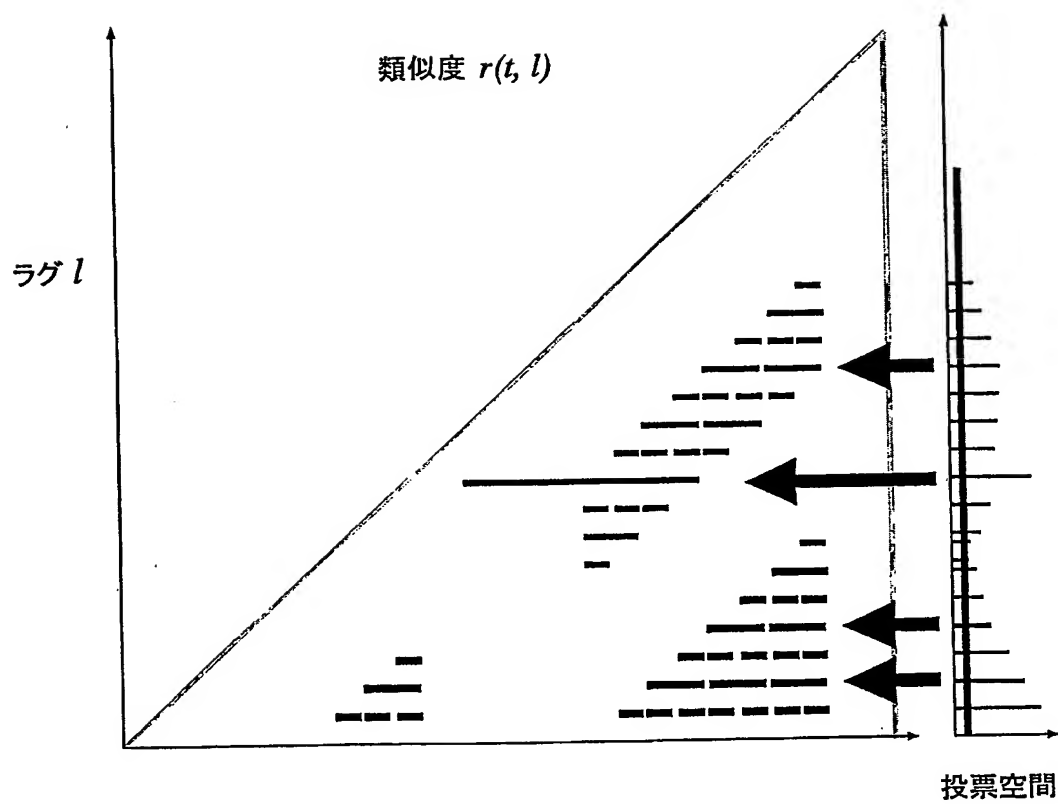


図31

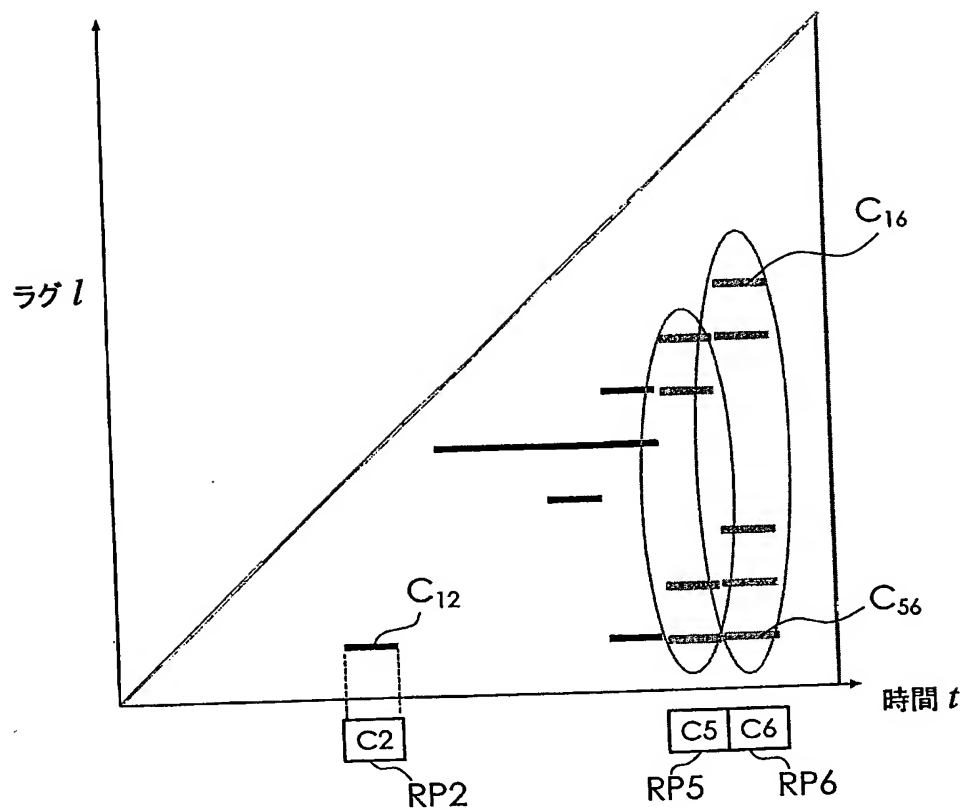


図32

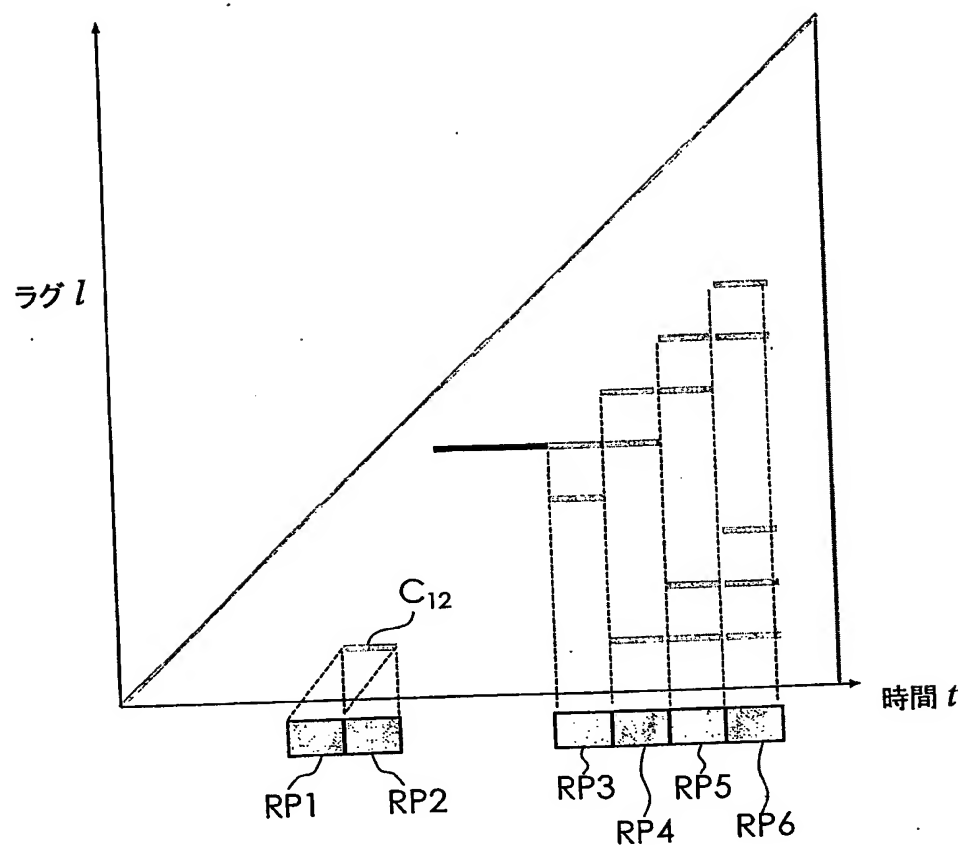


図33

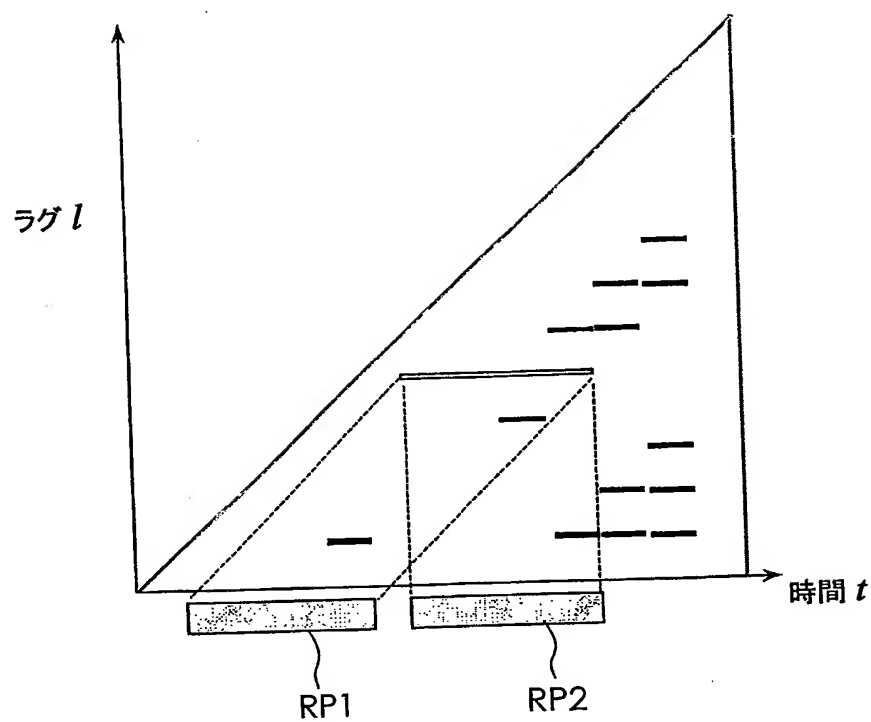
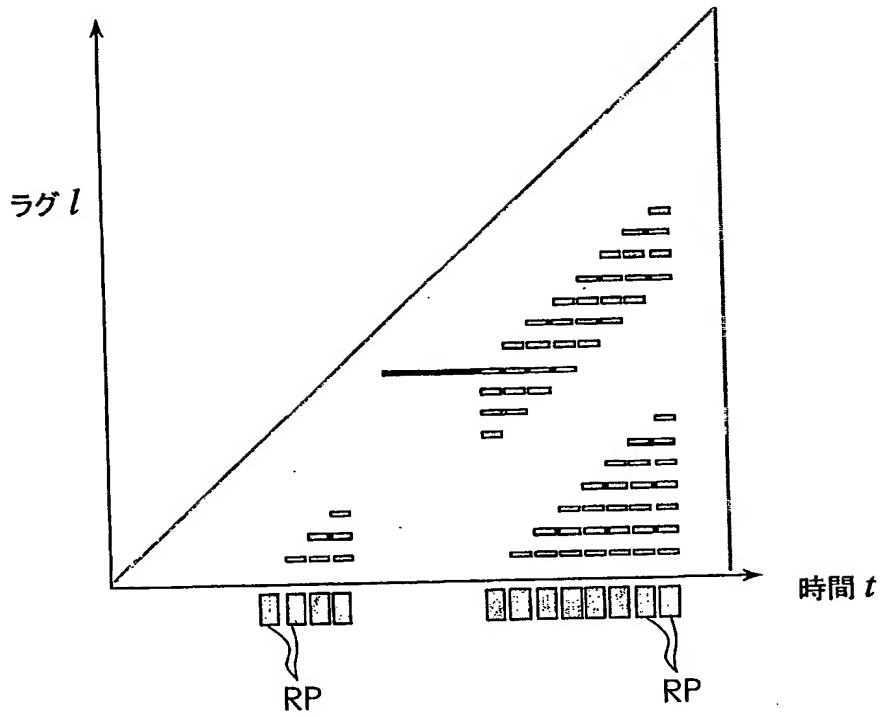


図34



35

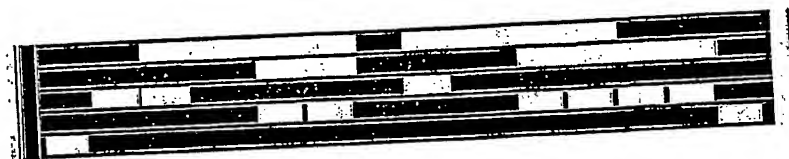




図36

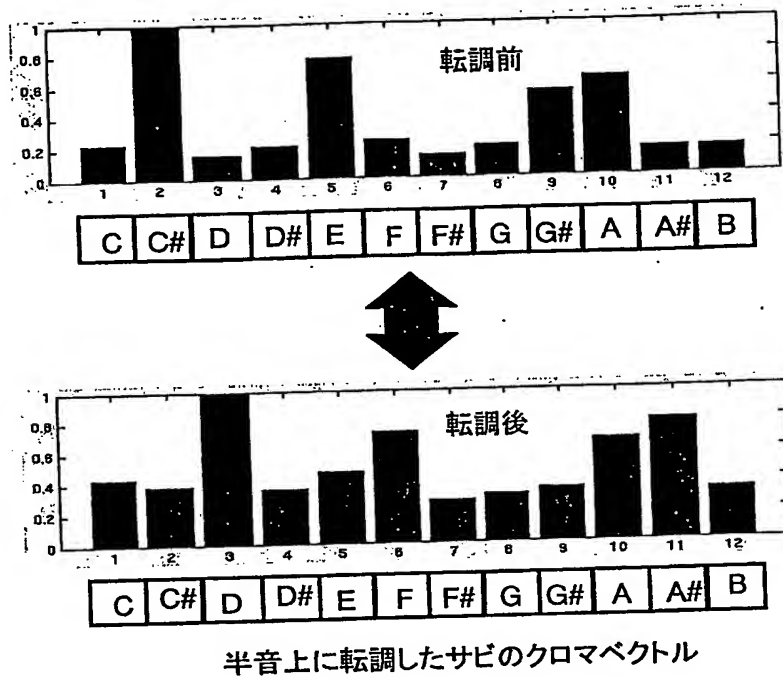
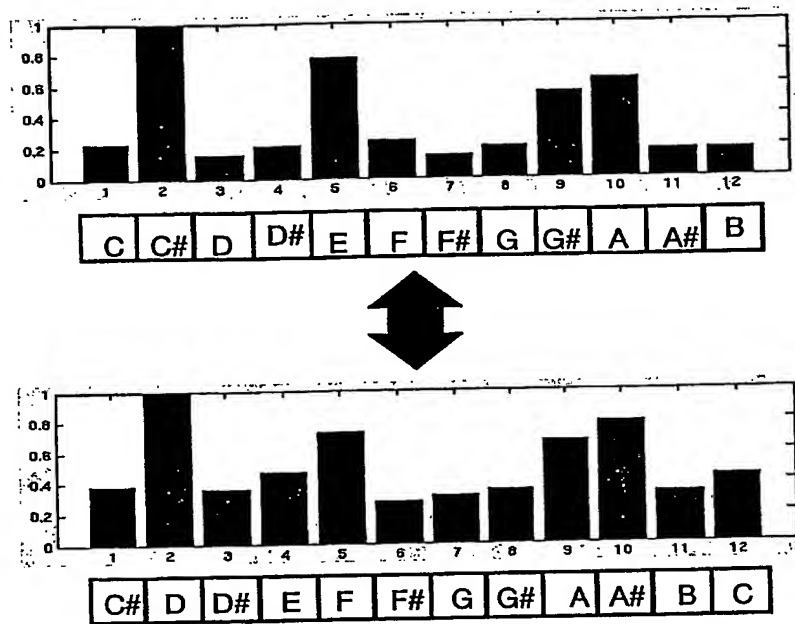
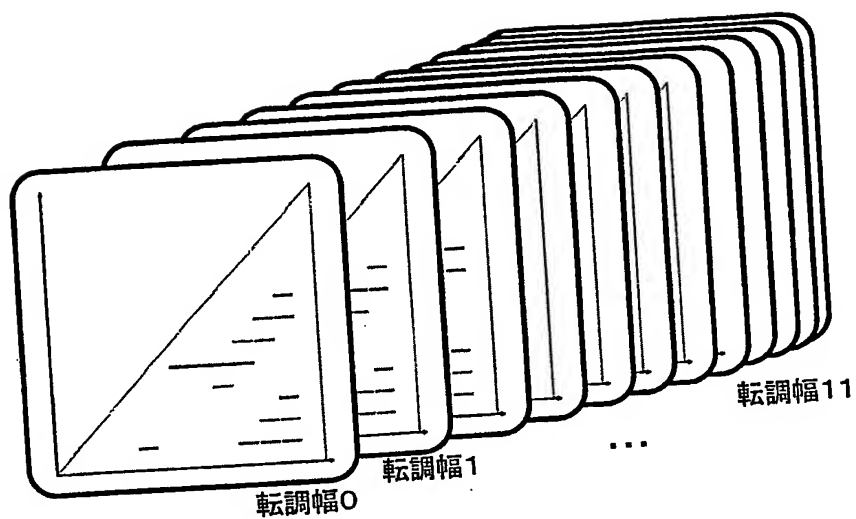


図37



転調幅1: 一つシフト

図38



類似線分のグループを統合

図39

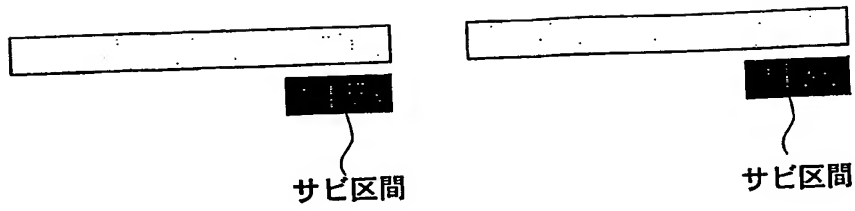
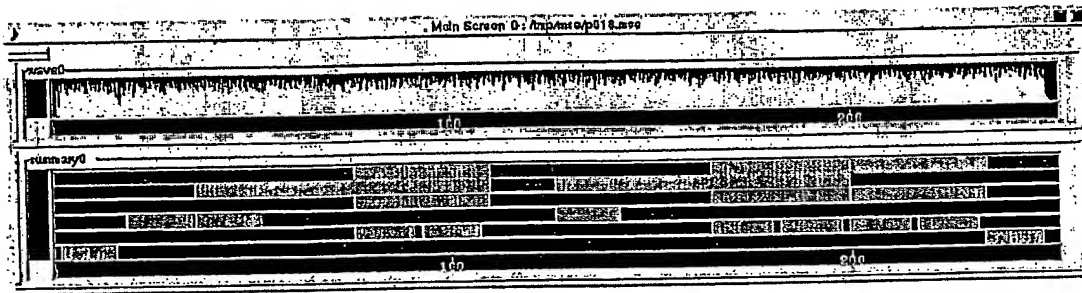


図40



図41



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**